PB82-224114

Correlations Between Age-Adjusted Mortality Rates for White Males and Females in the United States by County, 1968-1972

(U.S.) Environmental Protection Agency
Washington, DC

May 82

U.S. Department of Commerce
National Technical Information Service
NTIS

PB82-224114

EPA-600/4-80-026
May 1982

CORRELATIONS BETWEEN AGE-ADJUSTED MORTALITY RATES FOR WHITE MALES AND FEMALES IN THE UNITED STATES BY COUNTY, 1968-1972

by

Lance A. Wallace
and
Valarie J. Gill

Quality Assurance and Monitoring Systems Division

OFFICE OF MONITORING SYSTEMS AND QUALITY ASSURANCE
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENIVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA 22161

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA-600/4-80-026	2. ORD Report
3. RECIPIENT'S ACCESSION NO. PB82-224114	
4. TITLE AND SUBTITLE: Correlations Between-Age-Adjusted Mortality Rates for White Males and Females in the United States, By County: 1968 1972	5. REPORT DATE: MAY 1982
	6. PERFORMING ORGANIZATION CODE: ORD/OMTS/QAMSD (RD-680)
7. AUTHOR(S): Dr. Lance Wallace and Valarie J. Gill	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS: U.S. Environmental Protection Agency, Office of Research & Development, Office of Monitoring Systems & Quality Assurance, 401 M St., S.W., Washington D.C. 20460	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO.: In-House
12. SPONSORING AGENCY NAME AND ADDRESS: Same as #9	13. TYPE OF REPORT AND PERIOD COVERED: Technical - 1968-1972
	14. SPONSORING AGENCY CODE: EPA/600/19
15. SUPPLEMENTARY NOTES	

16. ABSTRACT

Intercorrelations among county mortality rates for about 50 causes of death were investigated for white males (WM) and white females (WF) for the 5-year period between 1968 and 1972. All possible pairwise correlations (1128 for WM and 1275 for WF) were calculated; those correlations significant at the $p < .0001$ level are listed. The geographic variations of the 20 strongest correlations for each sex are also listed. Most of the strongest correlations involve diseases that are closely associated with urban or rural factors. However diseases that are often grouped in epidemiological studies are shown to have negligible or even negative correlations; combining rates for such diseases could thus obscure rather than clarify possible environmental or other causal factors. The distribution of county mortality rates is shown to be neither normal nor log-normal; thus the interpretation of the correlation coefficient values is uncertain.

The UPGRADE computer system used to perform the calculations is shown to be useful, but also to lack certain features that would make such calculations efficient and cost effective. These features are to be added to the next release of UPGRADE.

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Correlation analysis, computer simulation, statistical analysis epidemiology, mortality rates geographic variations	Correlations Correlation Epidemiology Mortality	1201 1407 0605 0511

18. DISTRIBUTION STATEMENT: RELEASE TO PUBLIC	19. SECURITY CLASS *(This Report)*: UNCLASSIFIED	21. NO. OF PAGES: 76
	20. SECURITY CLASS *(This page)*: UNCLASSIFIED	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

DISCLAIMER

This report has been reviewed by the Office of Research and Development, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

FOREWORD

Monitoring activities provide the most important quantitative information available on the state-of-the-environment, thereby yielding a scientific measure of the impact of environmental regulatory actions. To generate valid monitoring data, research is necessary to evaluate and improve measuring instruments and techniques, develop criteria for designing monitoring networks, carry out effective statistical analyses of monitoring data, and give technical support to field monitoring activities. The Environmental Protection Agency's (EPA's) Office of Research and Development (ORD) carries forward a broadly based national research program in air pollution, water pollution, and other environmental topic areas. Within the ORD, the Office of Monitoring Systems and Quality Assurance oversees research and policies concerned with monitoring quality assurance, evaluation of measurement methods, application of advanced monitoring techniques, and development of improved methods for analyzing and handling environmental data.

This report is intended to assist researchers interested in mortality data by providing a complete set of the most significant correlations between all possible pairs of diseases among white males and females for the 1968-1972 period. These correlations may be used to suggest relationships that are worthy of further study, and to identify diseases that may be safely combined to provide larger numbers and more stable rates.

ABSTRACT

Intercorrelations among county mortality rates for about 50 causes of death were investigated for white males (WM) and white females (WF) for the 5-year period between 1968 and 1972. All possible pairwise correlations (1128 for WM and 1275 for WF) were calculated; those correlations significant at the $p \leq .0001$ level are listed. The geographic variations of the 20 strongest correlations for each sex are also listed. Most of the strongest correlations involve diseases that are closely associated with urban or rural factors. However, diseases that are often grouped in epidemiological studies are shown to have negligible or even negative correlations; combining rates for such diseases could thus obscure rather than clarify possible environmental or other causal factors. The distribution of county mortality rates is shown to be neither normal nor log-normal; thus the interpretation of the correlation coefficient values is uncertain.

The UPGRADE computer system used to perform the calculations is shown to be useful, but also to lack certain features that would make such calculations efficient and cost-effective. These features are to be added to the next release of UPGRADE.

CONTENTS

Foreword iii
Abstract iv
Figures vi
Tables vii
Acknowledgement viii

I. Introduction 1
II. Conclusions and Recommendations 2
III. Literature Review 4
IV. Methodology 5
V. Results. 7
VI. Discussion 9

References 12
Tables 14
Figures 32
Appendices 39

LIST OF FIGURES

Number		Page
1	Example of the reference scatterplots used to visually determine the range of rates for each disease variable . .	32
2	Example of individual scatterplots produced for each of the strongest 30 correlations for each sex	33
3	Frequency Distribution of Normalized Mortality Rates: Lung Cancer and Chronic Respiratory Disease	34
4	Cumulative Frequency Distribution of Mortality Rates . . .	35
5	Mortality Rates for Chronic and Acute Ischemic Heart Disease .	36
6	Cumulative Frequency Distribution of Mortality Rates for Lung Cancer and Chronic Respiratory Disease	37
7	Cumulative Frequency Distribution of Lung Cancer Mortality Rates - White Males	38

LIST OF TABLES

Number		Page
1	Correlation Coefficients for the Top Twenty (20) Correlations (White Males)	14
2	Correlation Coefficients for the Top Twenty (20) Correlations (White Females)	15
3	Disease Pairs Appearing in the Strongest 20 Correlations for Both White Males and Females	16
4	Disease Pairs with Similar Correlation Coefficients for Both Sexes	17
5	Correlation That are Strong for One Sex but Not the Other	18
6	Strongest Negative Correlations	19
7	Variation in Mortality Rate with County Population (Males)	20
8	Variation in Mortality Rate with County Population (Females)	21
9	Diseases Occuring Most Often in the Top Twenty Correlations	22
10	Comparison of Population-Weighted Correlations to Unweighted Correlations for 10 Selected Causes of Death (White Males)	23
11	Consensus of Like Correlations in Three Studies (Males)	24
12	Consensus of Like Correlations in Three Studies (Females)	25
13	Geographic Characterization for Cancer of Intestine vs. Cancer of the Rectum	26
14	Geographic Characterization for Major Cardiovascular Disease vs., Cancer All Forms	27
15	Geographic Characterization for Cancer of Rectum vs. Chronic Ischemic Heart Disease	28
16	Geographic Characterization or Chronic Ischemic Heart Disease vs. Cancer, All Forms	29

LIST OF TABLES CONTINUED

Number		Page
17	Geographic Characterization for Chronic Ischemic Heart Disease vs. Cirrhosis of the Liver	30
18	Geographic Characterization for Respiratory Cancer vs. Cirrhosis of the Liver	31

ACKNOWLEDGEMENT

We wish to thank both the President's Council on Environmental Quality (CEQ) for providing the UPGRADE computer system and the User Support Staff of Sigma Data Computing, Inc. for all of their assistance.

I. INTRODUCTION

The following study of correlations between county mortality rates was undertaken for two reasons:

1) To test the suitability of the UPGRADE (User-Prompted Graphics Data Evaluation) graphics and analysis computer system (1) for use in correlational studies of environmental health.

2) To identify causes of death with similarly varying county mortality rates. Causes with high correlations and similar etiologies may then be combined in further studies to obtain more stable rates. In addition, highly correlated pairs of diseases may also be mapped to determine their geographic patterns for each sex. Diseases with similar geographic patterns may then be studied further for possible geochemical or other environmental factors.

A knowledge of correlations between causes of death is useful for several reasons:

1) Strong correlations (or conversely, the lack of an expected strong correlation) may indicate relationships of mortality rates to one another, or to a common factor, that may be worthy of further study. For example, diseases that are often lumped together in epidemiological studies to get larger numbers and more stable rates, may not, in fact, be related strongly enough to justify such a practice.

2) Differences or similarities of correlations between sexes or races may indicate genetic or occupational factors of importance in disease etiology. For example, the lack of strong correlations in females to match the strong correlations for males between lung cancer and half a dozen other causes of death may indicate an occupational synergism occuring between these pairs of diseases for men.

3) Negative correlations may indicate "competing" causes of death, either in a real sense or in the sense that an inexperienced coroner may not be able to differentiate one specific cause of death from a related one.

II. CONCLUSIONS AND RECOMMENDATIONS

1. The correlation coefficients found in this study are reasonably dependable (i.e., reproducible to within 10-15% of their values under different selection of counties)--but only if great care is taken to remove outliers from the data base. For almost any given disease, 1-10 counties will have such extreme death rates as to cause major distortions in the Pearson correlation coefficient.

2. The strongest correlations are generally between diseases that are closely associated with county population density. This includes both diseases that increase with increasing population density (such as cirrhosis of the liver, chronic ischemic heart disease, and lung cancer) and those that decrease with increasing population density (such as acute ischemic heart disease and certain miscellaneous categories including other heart disease and symptoms, ill-defined).

3. On the other hand, many diseases that might be expected to have strong correlations on the basis of similar etiologies or similar bodily organs affected do not in fact show such correlations. Since such diseases are often grouped in epidemiological studies to provide larger numbers of deaths, the result of such grouping could well obscure rather than clarify relations between those diseases and possible environmental or other causal factors. Based on our findings, we are able to suggest categories of diseases that may safely be combined as well as those that cannot be combined for epidemiologic purposes. For example, it is clearly allowable to combine cancer of the large intestine with cancer of rectum to form colo-rectal cancer. On the other hand, for certain purposes one would be ill-advised to combine chronic heart disease with acute ischemic heart disease, or with other heart disease. Similarly, combining such diseases as cancer of the bladder, kidney, pancreas, liver, stomach, esophagus, etc. is not supported by the generally low correlations found among these diseases.

4. Diseases that are strongly correlated for one sex are usually strongly correlated for the other sex. However, two disease pairs appearing in the strongest 20 correlations for each sex were not strongly correlated for the other sex. Such diseases may be worthy of investigation for a possible genetic or cultural (e.g. occupational) link.

5. The distribution of county mortality rates is neither normal nor log-normal. Thus the interpretation of the Pearson correlation coefficients is uncertain. Because of the effect of outliers on the Pearson product-moment correlation coefficient r and also because r depends on normality assumptions that are overwhelmingly violated in this case of county

mortality rates, it would be valuable to employ nonparametric statistics such as the Spearman or Kendall measures. However, the computer cost of computing the Spearman coefficient is many times greater than for the Pearson coefficient when large number of data points are involved. Thus, the problem of employing nonparmetric statistics as a screening tool for identifying significant interdisease correlations has not yet been solved by the authors using the UPGRADE-SAS system.

6. Unweighted correlation coefficients were calculated throughout this study. An alternative option of weighting the mortality rates by county population was also investigated but not pursued. The latter option has the advantage of reducing the variance associated with the mortality rates of the small counties, but at the same time it masks the possible real differences in mortality rates between small and large counties.

7. Although the method of determining disease intercorrelations was probably successful in identifying most, if not all, of those diseases most strongly intercorrelated, it was not an efficient or cost-effective method. The authors were handicapped by not being able to create and save an improved data set for further analysis. This is a fault in the UPGRADE system that is being corrected. Until this feature is available, further work on this topic will be suspended.

III. LITERATURE REVIEW

Many epidemiological studies calculate correlation coefficents between mortality rates and environmental, demographic, and socioeconomic variables (2-5). Often rates for various diseases are grouped to obtain larger numbers of deaths and stabler rates. However, the basis for grouping these diseases often appears to be little more than a vague feeling that since they involve similar organs they must, therefore, be correlated. One purpose of this study is to determine those diseases that are, in fact, sufficiently strongly intercorrelated to justify grouping them. We are aware of several studies that attempt to carry out a similar program. One study by Sauer (6), using the same data base as the present study (NCHS mortality rates for 1968-1972), provides correlations of mortality rates for most major diseases calculated by state or by state economic area (SEA). A second study by Wellington, McDonald, and Wolf (7), gives state intercorrelations for many types of cancer, based on the 20-year cancer mortality rates between 1950 and 1969. The present study provides county intercorrelations and should thus be viewed as complementary to the other two studies. (See Appendices E,F,G, and H for a comparison of correlation coefficients determined in these two studies).

IV. METHODOLOGY

The UPGRADE computer system was used for all calculations in this study. UPGRADE is a system for analyzing data pertaining to the environment, natural resources, public health, and related topics. The system was developed by the President's Council on Environmental Quality (CEQ) and is partially supported by EPA. UPGRADE allows the user, through English language instructions, to perform both statistical and graphical analysis of several health, environmental, and demographic data bases.

The data for this report were obtained from UPGRADE's Integrated Data Base (IDB). This data base includes county level age-adjusted mortality and morbidity rates, along with census, socioeconomic and selected water quality data. Though UPGRADE's data has been age-adjusted, the crude mortality rates were orginally calculated by Herb Sauer of the University of Missouri, from the detailed mortality tapes provided by the National Center for Health Statistics (NCHS). The NCHS recorded every death between 1968-1972. The rates were calculated as follows:

$$\frac{\text{Total number of deaths collected}}{\text{1970 population} \times 4.5 \text{ years}}$$

Rates for 55 causes of death for four race-sex groups are contained in the data base. This study concentrates on disease-specific causes of death and, thus, employs only 51 of the 55 possible causes of death. Although all four race-sex groups were studied to some extent, only white males and females were studied exhautively. Whites were chosen rather than blacks for this pilot effort because of the far larger population base and correspondingly stable mortality rates. (For a list of the variables with corresponding ICD codes used in this study and in the studies mentioned in the Literature Review, see Appendices A and B).

Basic statistics for all 3082 U.S. counties were obtained for the 51 causes of death and reference scatterplots prepared for both sexes. (For an example see Figure 1). These graphs enabled us to visually determine the range of rates for each disease and remove obvious outliers. From these graphs, ranges for each cause of death were derived and recorded. In most cases, about 30 counties (1%) were identified as outliers for each disease variable. It should be noted that the graphs display only about 200 of the possible 3082 counties; nonetheless, it was believed and confirmed that a visual inspection of this subset of counties would be sufficient to determine the bounds containing nearly all the rates. Admittedly, some good values might have been thrown out, but it was felt that this process best eliminated the majority of obvious outliers. An alternative method would have been to examine each county rate to determine possible causes why it was an outlier, but this was neither feasible nor within the scope of this report.

Based on these set ranges, outliers for 20 variables at a time (the maximum allowed by the then-current version of UPGRADE) were eliminated. To intercorrelate all 51 variables, a series of 20 x 20 matrices were produced using the Statistical Analysis System (SAS) procedure. Some variation in execution was made between the males and females and will be discussed separately in Appendix I.

All of the matrices were examined for those intercorrelations with a p value $\leq .0001$. Out of the possible 1275 intercorrelations for white females, 181 exhibited a p value $\leq .0001$, while the males had 191 strong correlations out of a possible 1128 (see Appendices C & D).

Numbers of significant intercorrelations were reduced by omitting correlations between specific cancer sites and the variable "cancer, all forms and sites," and between specific cardiovascular diseases and the variable "major cardiovascular (CV) diseases." Thus, 152 correlations resulted for the females and 136 for the males. The top 30 of these correlations were selected and ranked for both sexes based on their correlation coefficient (r value) and their Student t value.

Individual scatterplots were produced for the top 30 ranked correlations for both sexes and examined for outliers. (See Figure 2 for an example.) If outliers were suspected, a new modified regression was run. This procedure resulted in certain changes in the order among the top correlations, but few sharp changes in correlation strengths. From these graphs, a final list of the top 20 correlations was created, based on the resulting r and t values.

Reliability of Results

After considerable running and rerunning of correlations, the authors have confidence that most of the correlations with $p \leq .0001$ are reproducible to within 10-15% of the r value, even if quite different selection procedures are employed to remove outliers. For example, Tables 1 and 2 show the r values obtained for the top 20 correlations for both males and females, as calculated from (1) the exclusion/filter procedure described above, and (2) the scatterplots method also already described. These two methods eliminated most, if not all, of the possible outliers in the data and thus assured the authors of a high level of confidence.

It should be noted, however, that for the less common causes of death, the numbers of deaths occurring even over 5 years in the less populated counties will be too small to provide a dependable death rate. This restriction may apply to as many as 1/4 to 1/2 of all U.S. counties (7).

V. RESULTS

The top 20 correlations for both white males and females are displayed in Tables 1 and 2. Eleven correlations appear in both tables (see Table 3). Of these eleven, four appear in each sex's top 10:

- Chronic ischemic heart disease vs. Cancer, all forms and sites
- Chronic ischemic vs. Cirrhosis of the liver
- Cancer of the intestine vs. Cancer of the rectum
- Other heart disease vs. Symptoms, ill-defined

Both diseases in the correlation--Other heart diseases vs. Symptoms, ill-defined--are "miscellaneous" categories that show strong increases in rural counties (9).

Nine correlations for each sex did not appear in the list of the top 20 for the opposite sex. Scatterplots were run of these correlations to compare to the r values already obtained for the opposite sex. Most of the correlations were relatively strong for both sexes (Table 4). However, two disease pairs for each sex had widely dissimilar correlation coefficients (Table 5).

Normality (or Lack Thereof) of the Mortality Rates

Use of the Pearson product-moment correlation coefficient assumes a normal distribution for each variable. How good is this assumption? We tested the distribution of county mortality rates for six causes of death for each of three race-sex groups. Not one of the 18 data sets passed chi-square tests for normality. Graphs of the distributions on standard Cartesian coordinates against the background of the normalized Gaussian curve showed distributions that were, in every case, more strongly clustered toward the mean and simultaneously more dispersed in the tails than the normal distribution. Such distributions are termed kurtic (more sharply peaked than the normal distribution). Graphs prepared on normal probability paper displayed similar behavior (Figure 3). The 18 distributions were then plotted on logarithmic probability paper but failed to display behavior asssociated with log-normal distributions (Figure 4-6).

The reasons for these results are not clear. Perhaps two or more distributions are, in fact, involved. For example, an urban-rural, high income-low income, educated-uneducated, or other spectrum of possibilities may be producing a bi- or multi-modal distribution. This hypothesis was tested using white male lung cancer mortality rates in 234 mostly urban U.S. counties; the result was a distribution that appeared

far closer to a normal distribution than that obtained from all 3082 U.S. counties (Figure 7). It is also possible that outliers (very high mortality rates arising from small county populations and low probabilities of death from some diseases) are producing a number of erroneously high death rates and corresponding inflated standard deviations for many diseases. These erroneously large standard deviations would then produce the observed apparent clustering of county mortality rates closer to the mean than otherwise expected. This hypothesis can be checked in two ways: by selecting only counties with large enough populations to reduce the probability of erroneously high mortality rates to near zero, or by selecting such common causes of death that the probability of erroneously high rates are greatly reduced. The latter approach has, in fact, already been followed, since the two combined-disease categories of "major cardiovascular diseases" and "cancer, all sites," have been graphed in Figure 4. The departure from log-normality is very evident. The former approach (setting a lower limit for county populations) suffers from the problem that omitting rural counties may, in fact, result in omitting mortality rates that are truly different (either higher or lower, depending on the disease) from rates in more populated counties. (See Figure 7) Thus, even if the distribution becomes more nearly normal or log-normal, it will be difficult to determine whether this is due to omitting erroneous rates or omitting rates that are truly different from those in the more populated counties.

Therefore, we are presently unable to identify a specific transformation of county mortality rates that will result in a normal distribution of these rates. Because of this problem, we are uncertain of the extent to which our results can be relied upon as quantitative estimates of the strengths of the correlations found. However, the relative strengths of the correlation coefficients may be more trustworthy.

Negative correlations are shown in Table 6. The two miscellaneous categories of "other heart disease" and "symptoms, ill-defined" appear nine and eight times respectively. Rheumatic heart disease, the third most frequently occurring cause of death in this table, appears only for white females. All five of the disease categories with which it has strong negative correlations have higher mortality rates in rural regions. However, for white males, no strong correlations with rheumatic heart disease are found. This phenomenon seems worthy of further study.

It should be remembered that the lack of normality of the county mortality rate distributions probably decreases the allowed range of the correlation coefficient in the negative direction. (For example, two log-normally distributed variables have a maximum negative r of -.369, although the positive limit remains at 1.0) (10). Thus, a negative coefficient may indicate a stronger relationship than a positive one of the same magnitude.

VI. DISCUSSION

The expected results of a study of correlations between disease-specific causes of death, would be a large number of strong correlations between those diseases involving similar body organs and/or etiologies. In fact, this was not the case. Out of the top 20 correlations for both white males and females, only a few involved similar body organs (see Results). Thus, a majority of correlations for each sex must be explained by some other causative factor. These could include such variables as water quality, air quality, income, education, smoking and drinking, and urban-rural residence.

Evidence that at least one of these factors did indeed influence the top 20 correlations can be found in a companion paper prepared by one of the authors (9). This paper studied correlations of mortality rates with county populations (i.e., the urban factor) for white males and females (see Tables 7 and 8). Six of the 48 causes of death investigated for white females showed a strong ($p \leq .0001$) increase in mortality rates in the more populous counties, while nine of 46 causes of death showed increases for males. The causes of death with rates that increased significantly with county population were:

Females	Males
° Cancer of the respiratory system	° Cancer of the esophagus
° Rheumatic heart disease	° Cancer of the intestine
° Chronic ischemic heart disease	° Cancer of the rectum
° Cirrhosis of the liver	° Cancer of the respiratory system
° Cancer of the breast	° Rheumatic heart disease
° Cancer, all forms and sites	° Aortic aneurysm
	° Cirrhosis of the liver
	° Cancer, all forms and sites

Four of these six diseases for white females and six of the nine for white males appear most often in the strongest 20 correlations for each sex (see Table 9). In fact, the "population-dependent" diseases noted above, appear 22 times for females and 25 times for males in the strongest 20 correlations for each sex.

Effect of Weighting Counties by Population

Several authors have suggested weighting the correlations by the county population (or the square root of the population). (11) To determine the magnitude of the effect this would have on the disease correlations, we selected ten diseases from among those appearing in the strongest 20

correlations for white males, weighted the rates by the 1970 county population and calculated their intercorrelations. (See Table 12 for results). Of those ten selected diseases, nine show significantly increasing rates with increasing county population (9). Thus weighting these diseases by county population would be expected to make the weighted correlations stronger than the unweighted values. This was, in fact, the case, as over 80% of the significant ($p \leq .0001$) correlations became stronger. Thus, it becomes clear that the population density difference contributes strongly to the intercorrelations among disease-specific causes of death in America. However, as is particularly well argued in Wellington, McDonald, and Wolf's book (8), the urban factor is probably a complex of more fundamental causative factors such as ethnicity, climate, diet, and pollution. Researchers should not remain content with using population density as an independent variable, but instead , strive to obtain significant results using variables more directly related to disease mortality.

Grouping Diseases on the Basis of Similar Body Sites

Many epidemiological studies assume strong intercorrelations among similar body organs and employ the practice of grouping these diseases together. This is done in order to produce larger numbers and, thus, more stable rates. For example, many studies combine acute ischemic, chronic ischemic, and other heart disease. However, our study shows very strong negative intercorrelations between chronic and other heart disease for both sexes. (This may be due to the fact that the two causes of death are apparently competitive for diagnosis. The miscellaneous "other" category often wins in rural areas, probably due to less sophistication in the coroners' training and experiences). Thus, combining these heart ailments to obtain larger numbers may obscure important relationships. (On the other hand, combining causes of death to reduce misclassification errors may be a legitimate practice). (12)

In another study (2), a group of diseases appear entitled "digestive-urinary cancer." This group included cancer of the esophagus, stomach, intestine, rectum, liver, pancreas, bladder, and kidney, presumably under the assumption that they are all closely associated. In fact, our findings indicate only a few of the disease rates to be strongly correlated. Examples include intestinal cancer with cancer of the rectum (ranked 5th and 9th for females and males respectively), and stomach cancer with rectal cancer (significant at $p \leq .0001$ for both sexes). On the other hand, many of these, such as cancer of the bladder and of the kidney, show no significant correlation whatsoever for either sex.

In another study (4), considerable effort went into grouping disease categories. As many as seven diseases were combined into one category; cancer of the esophagus, stomach, intestine, rectum, liver, pancreas, and other digestive cancers were grouped into the broad category of "digestive cancers". It should be noted that when correlations between mortality rates and a specific environmental factor are studie, parts of the anatomical system may be affected while others are not. For example, it has been noted

that lead in water supplies affects the stomach and intestines, but not the esophagus, yet all are part of the digestive system (13).

Of the studies referenced, the one most similar to ours was prepared by Herb Sauer of the University of Missouri (6). Sauer's study uses the same data (mortality data for 1968-1972) but groups it either by state or by state economic area (SEA) rather than by county. Thus, the two studies are complementary, and the findings of one should be in broad agreement with the other. In fact, this is the case, although we have often used somewhat different groupings of diseases, so that a detailed correspondence cannot be traced. However, Sauer's strongest correlation for white females are in substantial agreement with ours, as can be seen in Tables 11 and 12. For white females, both studies find the strongest relation between lung cancer and cirrhosis of the liver, yet another indication that personal habits (e.g. smoking and drinking) are perhaps the strongest determinants of chronic disease. However, Sauer finds a stronger relation between lung cancer and digestive cancers than we do, while we find relatively stronger relationships among some of the diseases of the heart or cerebrovascular system. The Wellington, McDonald, and Wolf book (8) has also taken a similar approach to correlating mortality rates, although only state rates were employed and only cancer was considered. Again, many similar results to those of Sauer and this paper were obtained.

Comparisons between the Sauer study, the work by Wellington, McDonald, and Wolf, and our report were made and a consensus of like correlations are displayed in Tables 11 and 12. Since different groupings of diseases were employed in each study, not all of the highly correlated disease pairs appear in these tables.

Geographic Variations

Geographic variations in disease rates may supply important clues to environmental or ethnic causes of disease. Atlases such as NCI's Cancer Atlas (14) and a forthcoming atlas of mortality from selected diseases (15) have been prepared or are being prepared to give researchers the necessary data for hypothesis generation. Similarly, the geographic pattern shown by a bivariate mapping of two diseases may suggest fruitful relationships to explore. We have therefore selected a number of pairs of highly correlated diseases from our "Top Twenty" lists for both white males and white females and prepared color maps of their variation.

The maps were prepared using the Domestic Information Display System (DIDS), a computerized mapping system developed at NASA-Greenbelt as a White House-sponsored interagency project. Rates for each disease were categorized into quartiles, and colors assigned to each of the 16 cells of the resulting 4x4 matrix. The resulting maps were studie, and regions of the country assigned to categories according to the predominant color, if any.

Geographic characterizations of six disease pairs showing high correlations for both white males & white females are presented in Tables 13-18.

REFERENCES

1. Users Guide for "CLIDE" (County Level Integrated Data Base for Epidemiology), Presidents Council on Environmental Quality, March 1978, Draft.

2. Sauer, H.I., L.E. Reed, "Minerals and the Risk of Chronic Diseases," Proceedings: Trace Substances in Environmental Health XII, University of Missouri, June 6-8, 1979.

3. Sauer, H.I. "Cause-Specific Death Rates and a Measure of Need." presented at the Data Use Conference, Dallas, Texas (NCHS), March 30, 1977.

4. Ericson, J.D. "Mortality in Selected Cities with Flouridated and Non-Flouridated Water Supplies," New England Journal of Medicine, 298: pp. 1112-1116, 1978.

5. Taves, Dr. R., Fluoridation and Mortality Due to Heart Disease," Nature, 272: pp. 361-362, 1978.

6. Sauer, H.I., "Geographic Patterns in the Risk of Dying and Associated Factors: U.S. 1968-1972," Contract No. HRA 230-76-0077, National Center for Health Statistics, DHEW.

7. Sauer, H.I., Personal Communication, Aug. 23, 1980.

8. Wellington, MacDonald and Wolf, "Cancer Mortality: Environmental and Ethnic Factors," Academic Press, New York, 1979.

9. Wallace, L.A., "Association of Mortality Rates with County Population," Unpublished.

10. DeVeaux, "Tight Upper and Lower Bounds for Correlation of Bivariate Distributions Arising in Air Pollution Modeling," Department of Statistics, Stanford University, Technical Report No. 5, November 11, 1976.

11. Fraumeni, J., Persons at High Risk of Cancer, Academic Press, New York, pp. 343-360, 1975.

12. Kleinman, J., Personal Communication, Aug. 11, 1980.

13. Berg, J.W., "Correlations Between Carcinogenic Trace Metals in Water Supplies and Cancer Mortality," Annals New York Academy of Sciences, 199: p. 249, 1972.

14. Mason, T.F. McKay, R. Hoover, W. Blot, and J. Fraumeni, Atlas of Cancer Mortality 1950-1969, National Cancer Institute, National Institutes of Health, DHEW Publication No. (HIH) 75-780, 1975.

15. Wallace, L.A. and V.J. Gill, National Atlas of Mortality Rates by County: 1968-1972," work in progress.

TABLE 1: CORRELATION COEFFICIENTS FOR THE TOP TWENTY (20) CORRELATIONS

White Males

Disease Title	Scatterplot Method	Exclusion/ Filter Method
1. Other Heart Disease - Symptoms, Ill-Defined	.286	.256
2. Cancer of the Respiratory System - Major Cardiovascular	.286	.302
3. Chronic Ischemic Heart Disease - Aortic Aneurysm	.268	.201
4. Chronic Ischemic - Cirrhosis of the Liver	.263	.221
5. Chronic Ischemic - Cancer, All Forms and Sites	.250	.217
6. Cirrhosis - Aortic Aneurysm	.246	.146
7. Cirrhosis - Cancer, All Forms	.243	.249
8. Cancer of the Respiratory System - Chronic Ischemic	.242	.223
9. Cancer of the Rectum - Cancer of the Intestine	.242	.168
10. Cancer of the Rectum - Chronic Ischemic	.241	.205
11. Major CV - Cancer, All Forms	.239	.220
12. Acute Ischemic Heart Disease - Cerebrovascular	.235	.301
13. Cancer of the Buccal Cavity, Pharynx - Cancer of the Respiratory System	.233	.170
14. Cancer of the Esophagus - Cirrhosis	.231	.186
15. Cancer of the Respiratory System - Chronic Respiratory	.231	.203
16. Cancer of the Respiratory System - Aortic Aneurysm	.226	.173
17. Aortic Aneurysm - Cancer, All Forms	.214	.178
18. Cancer of the Rectum - Cirrhosis	.205	.161
19. Cancer of the Respiratory System - Cancer Ill-Defined and Unspecified	.202	.175
20. Cancer of the Respiratory System - Cirrhosis	.201	.183

TABLE 2: CORRELATION COEFFICIENTS FOR THE TOP TWENTY (20) CORRELATIONS

White Females

Disease Title	Scatterplot Method	Exclusion/ Filter Method
1. Cancer of the Respiratory System - Cirrhosis	.224	.206
2. Cancer of the Intestine - Cancer of the Breast	.211	.161
3. Chronic Ischemic - Cancer, All Forms	.189	.238
4. Chronic Ischemic - Cirrhosis	.182	.198
5. Cancer of the Intestine - Cancer of the Rectum	.180	.145
6. Cancer of the Cervix - Major CV	.180	.189
7. Other Heart Disease - Symptoms, Ill-Defined	.179	.203
8. Rheumatic Heart Disease - Chronic Ischemic	.175	.193
9. Chronic Ischemic - Other Heart	-.172	-.201
10. Acute Ischemic - Cerebrovascular	.171	.173
11. Cancer of the Rectum - Rheumatic Heart	.170	.143
12. Aortic Aneurysm - Cirrhosis	.170	.152
13. Cancer of the Esophagus - Cirrhosis	.169	.118
14. Cancer of the Rectum - Chronic Ischemic	.167	.151
15. Rheumatic Heart - Cirrhosis	.166	.164
16. Diabetes - Major CV	.166	.210
17. Rheumatic Heart - Aortic Aneurysm	.161	.141
18. Cancer of the Rectum - Cancer of the Breast	.159	.155
19. Cirrhosis - Cancer, All Forms	.150	.194
20. Major CV - Cancer, All Forms	.146	.149

TABLE 3: DISEASE PAIRS APPEARING IN THE STRONGEST 20 CORRELATIONS FOR BOTH WHITE MALES AND WHITE FEMALES

Male Rank	Female Rank	Correlations
20	1	Cancer of the Respiratory System - Cirrhosis
5	3	Chronic Ischemic - Cancer, All Forms
4	4	Chronic Ischemic - Cirrhosis
9	5	Cancer of the Intestine - Cancer of the Rectum
1	7	Other Heart - Symptoms, Ill-Defined
12	10	Acute Ischemic - Cerebrovascular
6	12	Aortic Aneurysm - Cirrhosis
14	13	Cancer of the Esophagus - Cirrhosis
10	14	Cancer of the Rectum - Chronic Ischemic
7	19	Cirrhosis - Cancer, All Forms
11	20	Major CV - Cancer, All Forms

TABLE 4: DISEASE PAIRS WITH SIMILAR CORRELATION COEFFICIENTS FOR BOTH SEXES

A. Ranked White Female Correlation
With Corresponding White Male Rates

Rank (WF)	Correlations	Correlation Coefficient	
		Female	Male
8	Rheumatic Heart - Chronic Ischemic	.175	.170
9	Chronic Ischemic - Other Heart	-.172	-.197
15	Rheumatic Heart - Cirrhosis	.166	.139
17	Rheumatic Heart - Aortic Aneurysm	.161	.197

B. Ranked White Male Correlation
With Corresponding White Female Rates

Rank (WM)	Correlations	Correlation Coefficient	
		Male	Female
3	Chronic Ischemic - Aortic Aneurysm	.268	.136
8	Respiratory Cancer - Chronic Ischemic	.242	.134
15	Respiratory Cancer - Chronic Respiratory	.231	.119
16	Respiratory Cancer - Aortic Aneurysm	.226	.128
17	Aortic Aneurysm - Cancer, All Forms	.214	.113
18	Cancer of the Rectum - Cirrhosis	.205	.112
19	Respiratory Cancer - Cancer, Ill-Defined	.202	.102

TABLE 5: CORRELATIONS THAT ARE STRONG FOR ONE SEX BUT NOT THE OTHER

Rank (WF)		Correlation Coefficient WF	WM
11	Cancer of the Rectum - Rheumatic Heart Disease	.170	.130
16	Diabetes - Major CV Diseases	.166	.098
Rank (WM)			
1	Cancer of the Resp. System - Major CV Diseases	.084	.286
13	Cancer of the Resp. System - Ca. Buccal Cavity	.074	.233

TABLE 6: STRONGEST NEGATIVE CORRELATIONS

			WF	WM
Chronic Ischemic	vs.	Other Heart Disease	-.172	-.197
Ca. Rectum	vs.	Other Heart Disease	-.105	-.146
Aortic Aneurysm	vs.	Other Heart Disease	-.086	-.101
Ca. Breast	vs.	Other Heart Disease	-.137	NA
Ca. Intestine	vs.	Other Heart Disease		-.130
Acute Ischemic	vs.	Other Heart Disease		-.097
Ca. Intestine	vs.	Other Heart Disease	-.125	
Ca., All Forms	vs.	Other Heart Disease	-.123	
Acute Ischemic	vs.	Symptoms, Ill-Defined	-.118	-.196
Chronic Ischemic	vs.	Symptoms, Ill-Defined	-.137	-.155
Major CV	vs.	Symptoms, Ill-Defined	-.114	-.123
Ca. Rectum	vs.	Symptoms, Ill-Defined	-.083	-.125
Ca. All Forms	vs.	Symptoms, Ill-Defined	-.139	
Ca. Breast	vs.	Symptoms, Ill-Defined	-.122	
Ca. Intestine	vs.	Symptoms, Ill-Defined	-.123	
Other Heart Disease	vs.	Rheumatic Heart Disease	-.141	
Infections of Breast	vs.	Rheumatic Heart Disease	-.137	
Symptoms, Ill-Defined	vs.	Rheumatic Heart Disease	-.125	
Acute Ischemic	vs.	Rheumatic Heart Disease	-.118	
Cerebrobascular	vs.	Rheumatic Heart Disease	-.092	
Chronic Isch. Heart Disease	vs.	Cerebrovascular	-.141	-.093
Ca. Rectum	vs.	Cerebrovascular		-.082

Blanks indicate correlations that were not significant at $p \leq .0001$ for the particular sex.

Table 7. VARIATION IN MORTALITY RATE WITH COUNTY POPULATION
Age-Adjusted Mortality Rate per Million at Risk (1968-72)–White Males

Cause of Death	1970 White Male County Population (in thousands)					P*
	0-5	5-10	10-25	25-100	>100	
Tuberculosis, All Forms	25	30	32	31	35	–
Other Infective Disease	64	65	56	56	53	–
CA Buccal Cavity, Pharynx	45	44	47	54	64	.02
Cancer of Esophagus	26	29	33	38	46	.0001
Cancer of Stomach	94	95	92	92	107	–
Cancer of Intestine	152	157	162	175	208	.0001
Cancer of Rectum	39	46	56	62	72	.0001
Cancer of Liver, Gall B., Ducts	31	28	28	29	35	–
Cancer of Pancreas	110	105	108	108	111	–
Other Digestive Cancer	7	8	8	7	8	–
Cancer of Resp. System	512	529	566	600	645	.0001
Cancer of Breast	3	2	3	3	3	–
Cancer of Prostate	202	195	200	200	198	–
Cancer of Bladder	61	55	59	70	78	–
Cancer of Kidney, Etc.	44	43	43	47	46	–
Cancer of CNS	47	47	49	47	51	–
Residual Cancer	77	77	72	70	68	–
Lymphosarcoma, Etc.	38	36	40	41	43	–
Cancer Ill-Def. and Sec.	109	109	111	113	118	–
Hodgkins' Disease	22	22	21	20	21	–
Multiple Myeloma	24	24	24	25	24	–
Leukemia	108	96	97	92	91	–
Other Lymphatic	25	26	26	25	24	–
Neoplasms, Benign and Unspec.	23	20	22	25	26	–
Diabetes	170	166	171	174	173	–
Alcoholism	24	26	27	25	26	–
Rheumatic Heart Disease	55	61	65	72	83	.0001
Hypertension	117	133	129	121	104	–
Acute Ischemic Heart Dis	3006	3004	3019	2860	2629	.0001
Chronic Ischemic Heart	1223	1300	1445	1682	1930	.0001
Other Heart Disease	376	321	293	247	190	.0001
Cerebrovascular Disease	1207	1229	1252	1166	1056	.0003
Arteriosclerosis	198	198	202	199	180	–
Aortic Aneurysm	90	92	102	125	129	.0001
Influenza and Pneumonia	415	403	397	387	393	–
Chronic Resp. Disease	403	376	395	423	397	–
Cirrhosis of Liver	115	116	128	158	219	.0001
Chronic Nephritis	44	45	43	47	35	–
Infections of Kidney	50	46	49	44	39	–
Congenital Heart & Cir.	42	44	44	42	41	–
Other Congenital	48	45	47	45	42	–
Other Early Infancy	237	234	228	216	199	.02
Major Cardiovascular Diseases	6355	6416	6587	6549	6311	–
Cancer, All Sites and Forms	1,777	1,774	1,848	1,926	2,071	.0001

*Probability that the increase (decrease) in rates is due to chance (Pearson product-moment correlations applied to all counties)

Table 8. VARIATION IN MORTALITY RATE WITH COUNTY POPULATION
Age-Adjusted Mortality Rate per Million at Risk (1968-72)—White Females

Cause of Death	1970 White Female County Population (in thousands)					P*
	0-5	5-10	10-25	25-100	>100	
Tuberculosis, All Forms	9	10	10	10	10	–
Other Infective Disease	56	52	46	42	40	.03
Cancer of Buccal Cavity, Pharynx	16	15	16	17	19	–
Cancer of Esophagus	9	8	9	9	12	.03
Cancer of Stomach	50	46	48	45	52	–
Cancer of Intestine	143	147	156	160	173	.002
Cancer of Rectum	28	29	34	38	40	.002
Cancer of Liver, Gall B., Ducts	37	31	31	32	32	–
Cancer of Pancreas	61	62	64	64	66	–
Other Digestive Cancer	8	6	6	6	6	–
Cancer of Resp. System	85	90	97	107	126	.0001
Cancer of Breast	212	219	228	249	279	.0001
Cancer of Cervix	52	59	61	61	50	–
Cancer of Uterus	42	46	46	44	46	–
Other Female Cancer	81	84	91	95	99	.007
Cancer of Bladder	20	17	19	22	23	–
Cancer of Kidney	22	21	21	21	21	–
Cancer of CNS	31	32	33	30	32	–
Residual Cancer	46	44	43	40	39	–
Lymphosarcoma, Etc.	23	24	26	28	28	–
Cancer, Ill-Def. and Sec.	85	97	91	89	92	–
Hodgkins Disease	11	12	11	12	13	–
Multiple Myeloma	17	15	16	17	17	–
Leukemia	58	59	59	56	57	–
Other Lymphatic	16	16	16	15	16	–
Neoplasms, Benign and Unspec.	20	21	18	20	21	–
Diabetes	192	180	187	186	175	–
Alcoholism	6	5	5	6	7	–
Rheumatic Heart Disease	51	50	55	67	82	.0001
Hypertension	104	111	112	97	86	.05
Acute Ischemic Heart Disease	1,197	1,212	1,202	1,161	1,114	.01
Chronic Ischemic Heart	900	938	1,026	1,179	1,263	.0001
Other Heart Disease	214	195	175	148	137	.003
Cerebrovascular Disease	1,000	978	996	962	899	.009
Arteriosclerosis	162	162	164	168	149	–
Aortic Aneurysm	28	26	28	33	34	–
Influenza and Pneumonia	264	252	243	233	225	.04
Chronic Resp. Disease	89	88	89	95	94	–
Cirrhosis of Liver	51	50	57	74	101	.0001
Chronic Nephritis, Etc.	28	29	29	25	22	–
Infections of Kidney	47	41	42	37	34	–
Congenital Heart and Circ.	33	35	35	34	33	–
Other Congenital	43	41	45	43	40	–
Other Early Infancy	172	161	164	154	143	.04
Major Cardiovascular Diseases	3,712	3,723	3,813	3,869	3,815	–
Cancer, All Sites	1,153	1,178	1,227	1,254	1,359	.0001

*See note to Table 7.

TABLE 9: DISEASES OCCURRING MOST OFTEN IN THE TOP TWENTY CORRELATIONS

FEMALE			MALE
Cirrhosis	6	7	Ca. Respiratory System
Chronic Ischemic	5	6	Cirrhosis
Ca. Rectum	4	5	Chronic Ischemic
Rheu. Heart	4	4	Aortic Aneurysm
Major CV	3	4	Ca. All Forms
Ca. All Forms	3	3	Ca. Rectum
Ca. Intestine	2	2	Major CV
Ca. Breast	2	1	Other Heart
Other Heart	2	1	Symptoms, Ill-Defined
Aortic Aneurysm	2	1	Ca. Intestine
Ca. Respiratory System	1	1	Acute Ischemic
Ca. Cervix	1	1	Cerebrovascular
Symtoms, Ill-Defined	1	1	Ca. Buccal Cavity
Acute Ischemic	1	1	Ca. Esophagus
Cerebrovascular	1	1	Chronic Resp.
Ca. Esophagus	1	1	Ca. Ill-Defined & Sec.
Diabetes	1		

TABLE 10: COMPARISON OF POPULATION-WEIGHTED CORRELATIONS TO UNWEIGHTED CORRELATIONS FOR 10 SELECTED CAUSES OF DEATH (WHITE MALES)

Disease Pairs			Weighted r*	Unweighted r**
Cirrhosis	vs	Ca. Resp. System	.251	.178
		Chronic Ischemic	.539	.222
		Ca. Rectum	.311	.168
		Rheumatic Heart	.235	.093
		Aortic Aneurysm	.169	.146
		Major CV	.081	.081
		Ca. All Forms	.415	.249
		Ca. Intestine	.301	.127
		Other Heart Disease	-.214	NS
Ca. Resp. System	vs	Chronic Ischemic	.280	.220
		Ca. Rectum	.079	NS
		Aortic Aneurysm	.177	.173
		Major CV	.299	.302
		Ca. Intestine	.167	.096
Chronic Ischemic	vs	Ca. Rectum	.406	.204
		Aortic Aneurysm	.137	.201
		Ca. All Forms	.477	.217
		Ca. Intestine	.392	.128
		Other Heart Disease	-.272	-.214
		Rheumatic Heart	.210	.158
Ca. Rectum	vs	Rheumatic Heart	.222	.130
		Aortic Aneurysm	.092	.091
		Major CV	.149	NS
		Ca. Intestine	.423	.168
		Other Heart	-.215	-.142

Disease Pairs			Weighted r*	Unweighted r**
Rheumatic Heart	vs	Aortic Aneurysm	.216	.118
		Ca. All Forms	.096	NS
		Ca Intestine	.175	.074
		Other Heart Disease	-.241	-.086
Aortic Aneurysm	vs	Ca. All Forms	.195	.178
		Ca. Intestine	.102	.077
		Other Heart Disease	-.193	-.101
Major CV	vs	Ca. All Forms	.327	.227
		Ca. Intestine	.120	NS
Ca., All Forms	vs	Other Heart	-.096	NS
Ca. Intestine	vs	Other Heart	-.227	-.180

NS - Not Significant at $p \leq .0001$

* Correlation coefficient calculated by multiplying county mortality rate by 1970 county population.

** Correlation coefficients in this column differ somewhat from those in Table 1 due to different criteria for removing outliers (See Appendix I).

TABLE 11: RANKED CONSENSUS OF LIKE CORRELATIONS

Males

	Wallace-Gill (Counties)	Sauer (SEA)	Wellington-McDonald-Wolf (States)
Ca. Resp. - Major CV	2	2	
Cirrhosis - Ca. All Forms	7	11	
Ca., Rectum - Intestine	9		2
Ca. All Forms - Major CV	11	5	
Buccal Cavity - Ca. Resp.	13		14
Ca., Resp. - Ca. Ill-Defined	19	6	
Ca. Ill-Def. & Sec. - Cerebrovascular	.175	13	
Ca. Rectum - Bladder	.156		3
Ca. Ill-Def. & Sec. - Major CV	.152	7	
Ca. Resp. - Cerebrovascular	.141	14	
TB - Influenza & Pneumonia	.135	15	
Major CV - Influenza & Pneumonia	.119	9	
Ca. Ill-Def. & Sec. - TB	.117	19	
Ca. Esophagus - Ca. Rectum	.114		11
Ca. Intestine - Ca. Bladder	.110		4
Ca. Esophagus - Ca. Intestine	.086		5
Ca. Buccal - Ca. Esophagus	.075		7

Decimal numbers denote the correlation coefficient for the particular correlation, indicating it was not ranked in the top twenty but was significant at $p \leq .0001$. Blanks indicate correlation that were not ranked in the top twenty or did not correspond to the disease categories chosen by the investigator.

TABLE 12: RANKED CONSENSUS OF LIKE CORRELATIONS

Females

	Wallace-Gill (Counties)	Sauer (SEA)	Wellington-McDonald-Wolf (States)
Cirrhosis - Ca. Resp.	1	1	
Ca. Intestine - Ca. Breast	2		4
Ca. Intestine - Ca. Rectum	5		1
Major CV - Ca. Cervix	6	10	
Diabetes - Major CV	16	5	
Ca. Rectum - Ca. Breast	18		3
Ca. All Forms - Cirrhosis	19	15	
Ca. All Forms - Major CV	20	7	
Cirrhosis - Chronic Resp.	.130	17	
Ca. Intestine - Other Female Ca.	.125		5
Major CV - Ca. Ill-Def. & Sec.	.123	11	
Rheumatic Heart - Ca. Breast	.098	12	
Ca. Breast - Lymphosarcoma	.095		9
Ca. Breast - Other Female Ca.	.088		2
Ca. Rectum - Other Female Ca.	.081		7
Major CV - Influenza & Pneumonia	.080	16	

Decimal numbers denote the correlation coefficient for the particular correlation, indicating it was not ranked in the top twenty but was significant at $p \leq .0001$. Blanks indicate correlation that were not ranked in the top twenty or did not correspond to the disease categories chosen by the investigator.

TABLE 13: GEOGRAPHIC CHARACTERIZATION FOR CANCER OF THE INTESTINE VS CANCER OF THE RECTUM

Cancer of Intestine	Cancer of Rectum	Sex	Geographic Location
HIGH	HIGH	WM	Mid-Atlantic, New England, Eastern North Central through Minnesota, but not North and South Dakota
		WF	Mid-Atlantic, New England, North Cental through Illinois and Wisconsin, but not Minnesota or North and South Dakota
MEDIUM	MEDIUM	WM	Pacific Coast, Arizona, New Mexico
		WF	Pacific Coast, Minnesota, Iowa
LOW	LOW	WM	South, Sunbelt
		WF	South, Sunbelt (except Arizona)
MIXED	HIGH	WF	Arizona (northeast corner)

TABLE 14: MAJOR CARDIOVASCULAR DISEASES VS CANCER, ALL FORMS

Major Cardiovascular Diseases	Cancer, All Forms	Sex	Geographic Location
HIGH	HIGH	WM	Atlantic Coast from Maine to Georgia
		WF	Pennsylvania, New York, Southern Ohio, Maine
HIGH	MIXED	WM	Southest (east of Appalachia), Pennsylvania, Ohio
		WF	Southeast (east of Appalachia), Nevada, Northern Michigan
LOW	HIGH	WM	Florida, Texas Gulf, Connecticut, Montana
		WF	Illinois, Southern Michigan, Indiana, Wisconsin, Missouri
LOW	LOW	WM	Oregon, Idaho, Utah, Arizona, New Mexico, Colorado, Southwestern Texas
		WF	Alaska, Arizona, Eastern Oregon
MEDIUM	MEDIUM	WM	Illinois, Michigan
LOW	MIXED	WF	Florida, Mid-West (North and South Dakota, Minnesota Nebraska, Kansas, Arkansas, Texas, Southeastern New Mexico, Montana)
MIXED	HIGH	WM	Nevada, California, Louisiana

TABLE 15: CANCER OF THE RECTUM VS CHRONIC ISCHEMIC

Cancer of the Rectum	Chronic Ischemic	Sex	Geographic Location
HIGH	HIGH	WM	Mid-Atlantic, New England, Illinois, California, Ohio
		WF	Mid-Atlantic, New England
MIXED	HIGH	WM	South Carolina, Virginia, Atlantic Coast of North Carolina
		WF	Illinois, Wisconsin, Virginia, North and South Carolina, Southern Ohio, California (parts)
MEDIUM	MEDIUM	WM	Pacific Coast, Alaska, Alabama, Arizona
		WF	Pacific Coast, Arizona
MIXED	LOW	WM	West (from about 100th Meridian)
LOW	LOW	WF	Mid-West

TABLE 16: CHRONIC ISCHEMIC HEART DISEASE VS CANCER, ALL FORMS

Chronic Ischemic Heart Disease	Cancer, All Forms	Sex	Geographic Region
HIGH	HIGH	WM	New England*, Mid-Atlantic**, Southeast Louisiana
		WF	New England, Mid-Atlantic, Louisiana
HIGH	MIXED	WM	South Carolina, Virginia, Illinois, California
		WF	California, South Carolina
LOW	MIXED	WM	Texas, Oklahoma, Kansas, Missouri, Arkansas
		WF	Appalachia, Texas, Oklahoma, Arkansas (mostly everything west of Mississippi to the Cascades and Sierra Nevadas)
HIGH	LOW	WF	Southern Atlantic***
MIXED	LOW	WM	Kentucky
LOW	HIGH	WF	Montana
LOW	LOW	WM	Western Texas, New Mexico, Colorado, Utah, Idaho, Wyoming
MEDIUM	MEDIUM	WF	Pacific Coast, North Central****

*Particularly VERMONT, NEW HAMPSHIRE, WESTERN MAINE
**Particularly PENNSYLVANIA, NEW YORK, NEW JERSEY, MARYLAND, CONNECTICUT, MASSACHUSETTS, DELAWARE, SOUTHERN OHIO
***Particularly VIRGINIA, NORTH CAROLINA, GEORGIA
****Particularly MINNESOTA, WISCONSIN, MICHIGAN, INDIANA

TABLE 17: CHRONIC ISCHEMIC HEART DISEASE VS CIRRHOSIS OF LIVER

Chronic Ischemic Heart Disease	Cirrhosis of Liver	Sex	Geographic Location
HIGH	HIGH	WM	Mid-Atlantic, California
		WF	Mid-Atlantic, New England, California
HIGH	MIXED	WM	Southeastern South Carolina, Virginia, Illinois
		WF	South Carolina, Southern Ohio, Pennsylvania, Virginia, Illinois
MIXED	HIGH	WM	New Mexico, Wyoming, Colorado, Southern Florida
		WF	Arizona, Nevada, New Mexico
LOW	HIGH	WM	Alaska, Southern Texas
		WF	Southern Florida
LOW	LOW-MIXED	WM	Texas, Arkansas, Oklahoma, Nebraska, Kansas, North Dakota, South Dakota, Minnesota, Iowa, Missouri
MEDIUM	MEDIUM	WM	Appalachia, Michigan, Indiana, Wisconsin, Iowa
		WF	Michigan, Wisconsin, Indiana, Northern Ohio, Alabama
LOW	MIXED	WF	Texas, Oklahoma, Arkansas, Northern Louisiana, Mississippi, Alabama
LOW	LOW	WF	Montana, South Dakota
HIGH	LOW	WF	Kentucky, Eastern Colorado

TABLE 18: RESPIRATORY CANCER VS CIRRHOSIS OF LIVER

Respiratory Cancer	Cirrhosis of Liver	Sex	Geographic Location
HIGH	HIGH	WM	New England, California, Florida
		WF	New England*, California, Florida, Nevada, Arizona, New Mexico, Washington (Seattle-Tacoma-Everett), Gulf Coast, Alaska
LOW	LOW	WM	Tennessee, Kentucky, Virginia
		WF	West**, Southeast
HIGH	MIXED	WM	Georgia, South Carolina, Lower Mississippi River
LOW	HIGH	WM	West, Southwest***

*Particularly CONNECTICUT, MASSACHUSETTS, SOUTHERN VERMONT AND NEW HAMPSHIRE, EASTERN NEW YORK STATE, COASTAL PARTS OF MAINE, MOST OF NEW JERSEY

**Particularly WESTERN PORTION OF NORTH AND SOUTH DAKOTA, NEBRASKA, KANSAS, SOUTHERN PORTION OF MONTANA, IDAHO, UTAH

***Particularly NEW MEXICO, COLORADO, WYOMING

Figure 1

Example of the Reference Scatterplots Used to Visually Determine the Range of Rates for Each Disease Variable

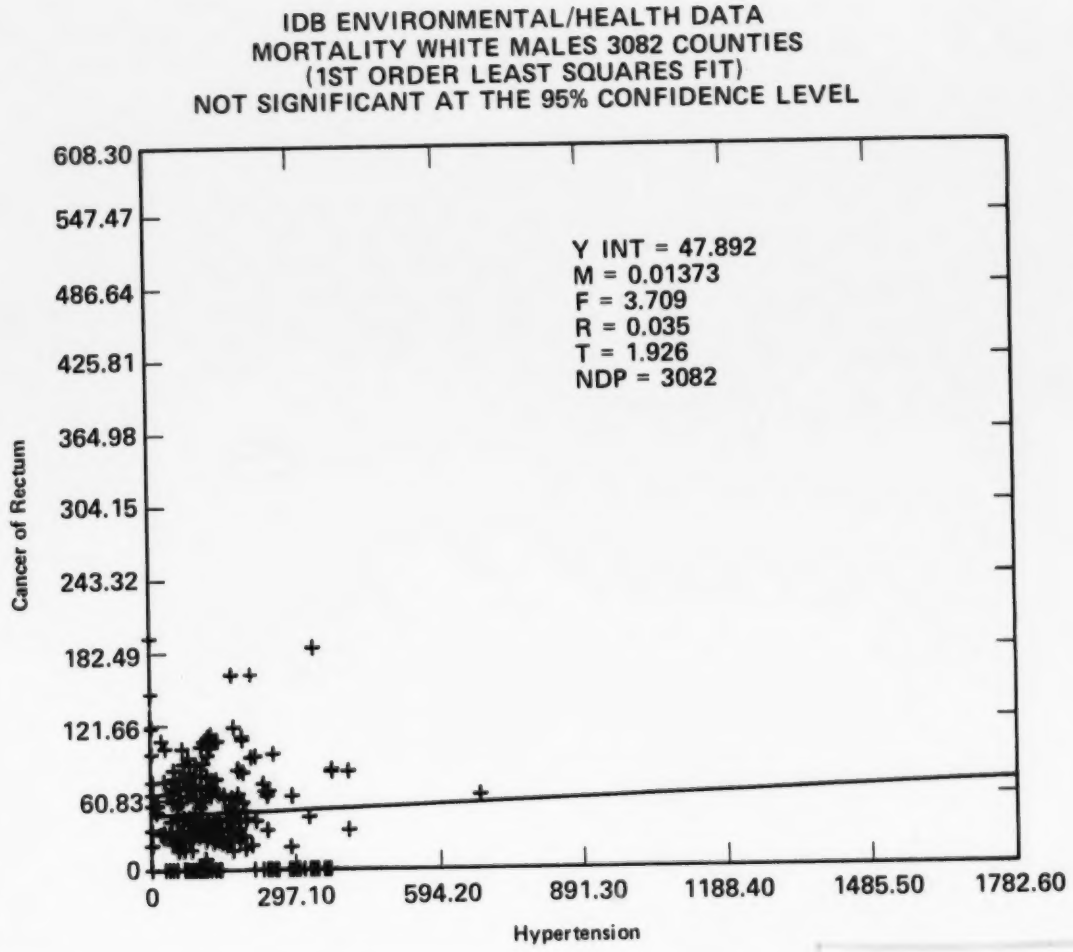


Figure 2
Example of Individual Scatterplots Produced for Each of the Strongest 30 Correlations for Each Sex

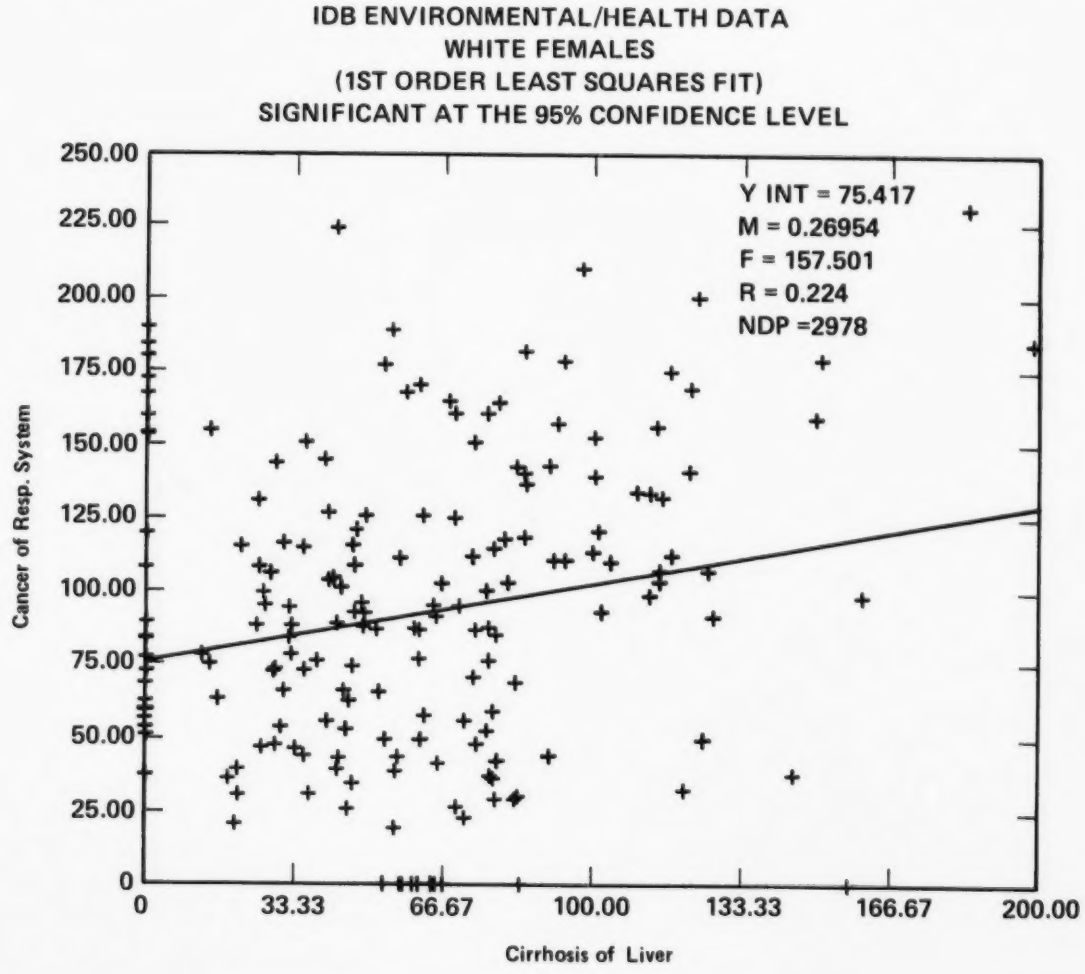


Figure 3

Frequency Distribution of Normalized Mortality Rates: Lung Cancer and Chronic Respiratory Disease (1968-72)

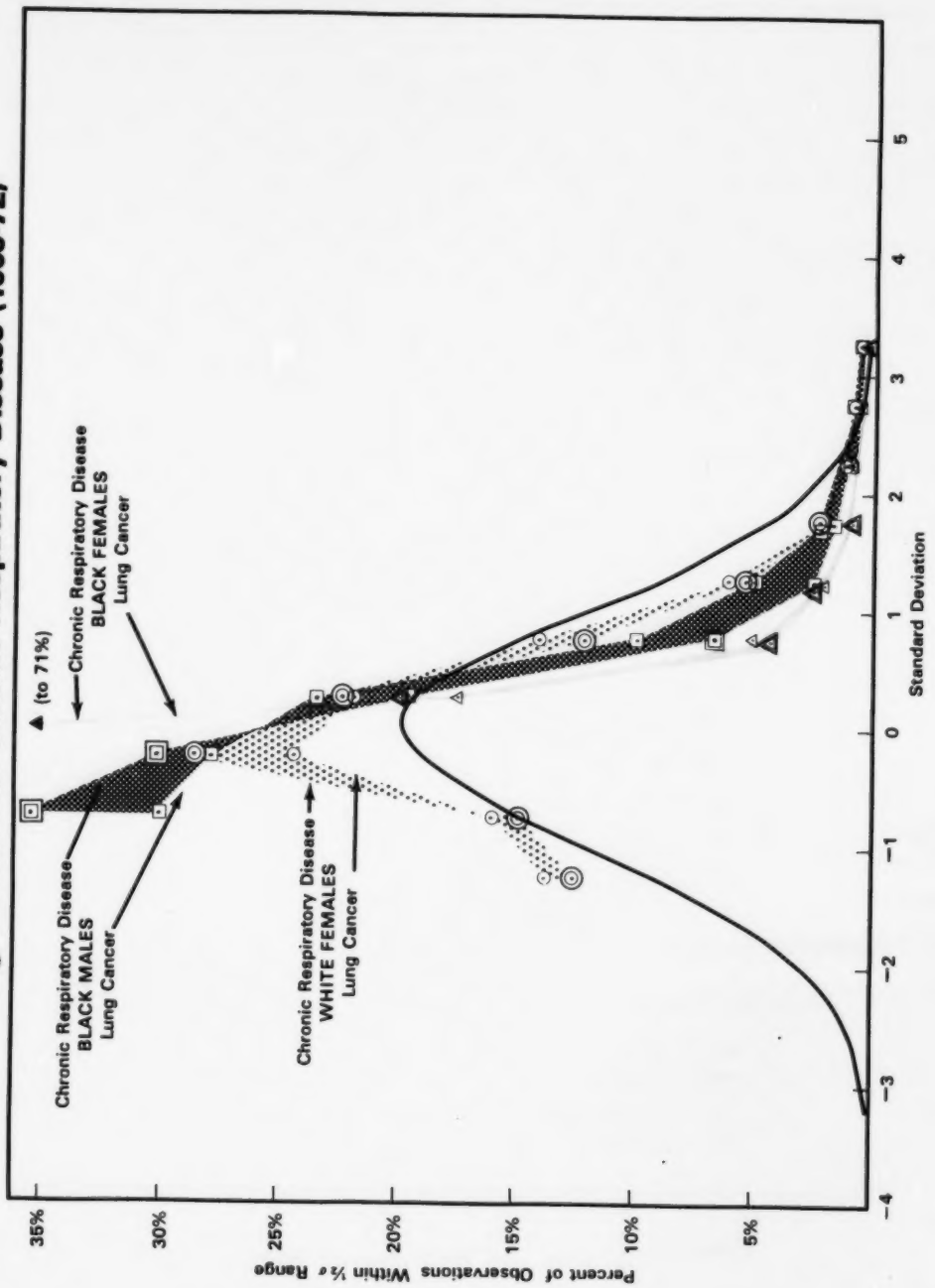


Figure 4
Cumulative Frequency Distribution of Mortality Rates: (1968-72)

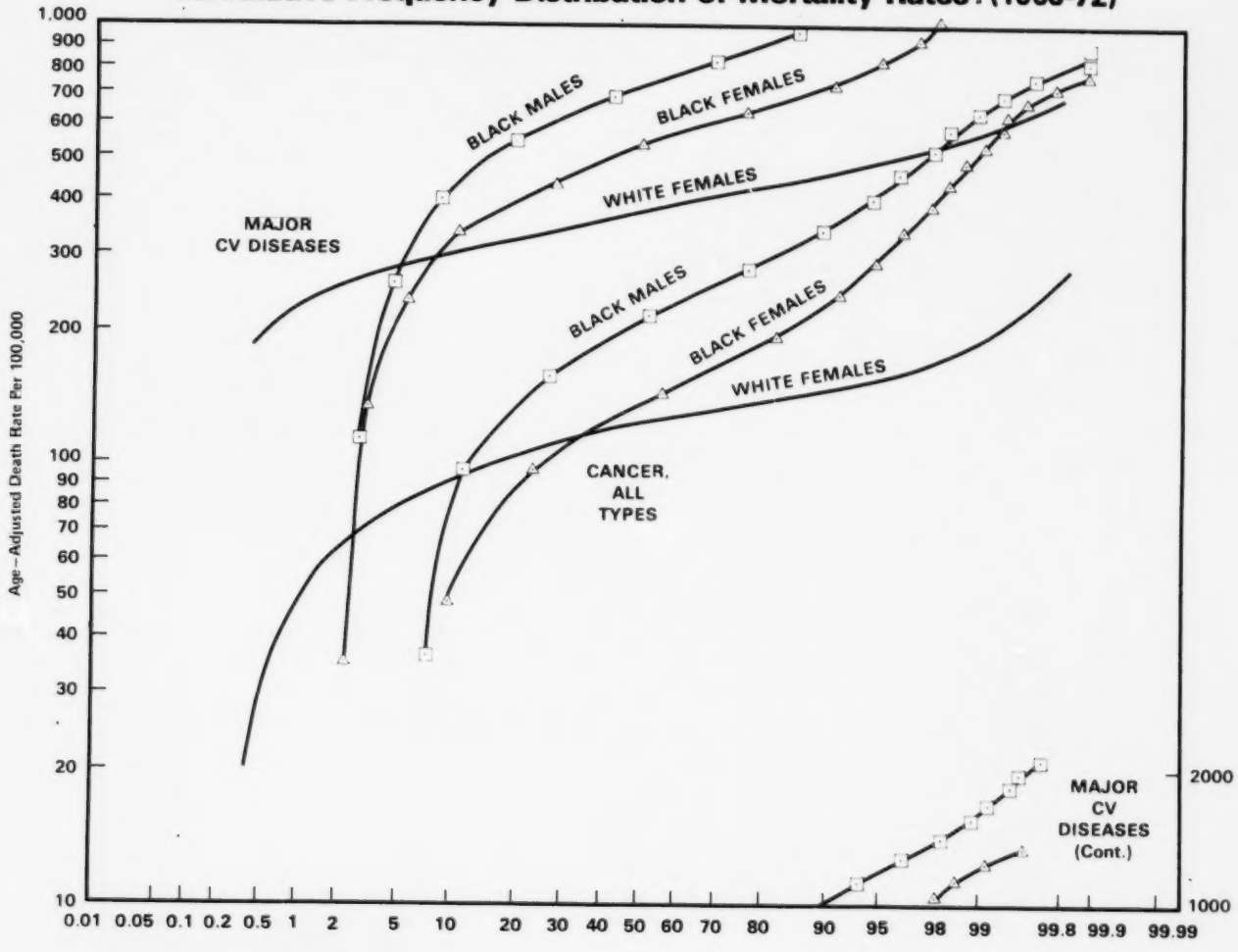


Figure 5

Mortality Rates For Chronic and Acute Ischemic Heart Disease

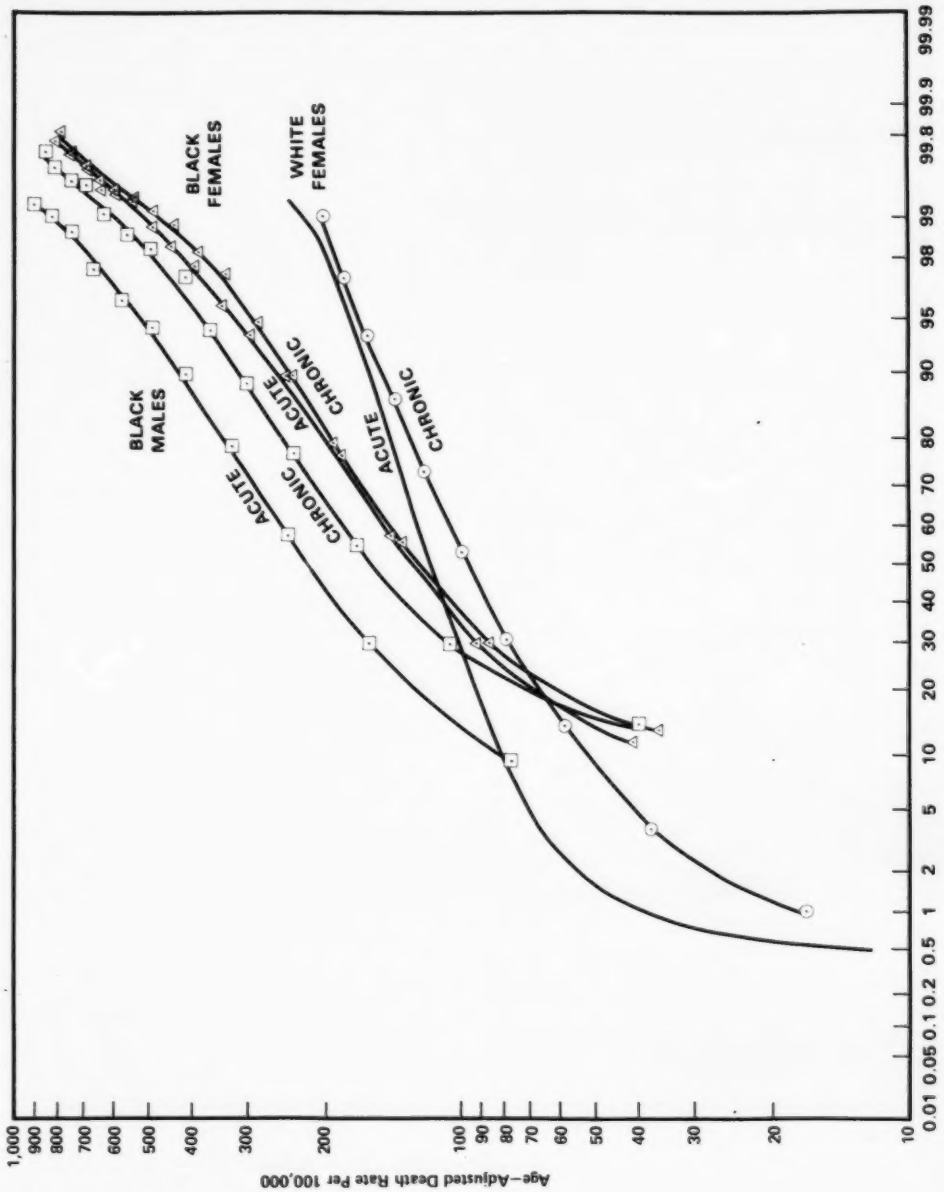


Figure 6
Cumulative Frequency Distribution of Mortality Rates
For Lung Cancer and Chronic Respiratory Disease (1968-72)

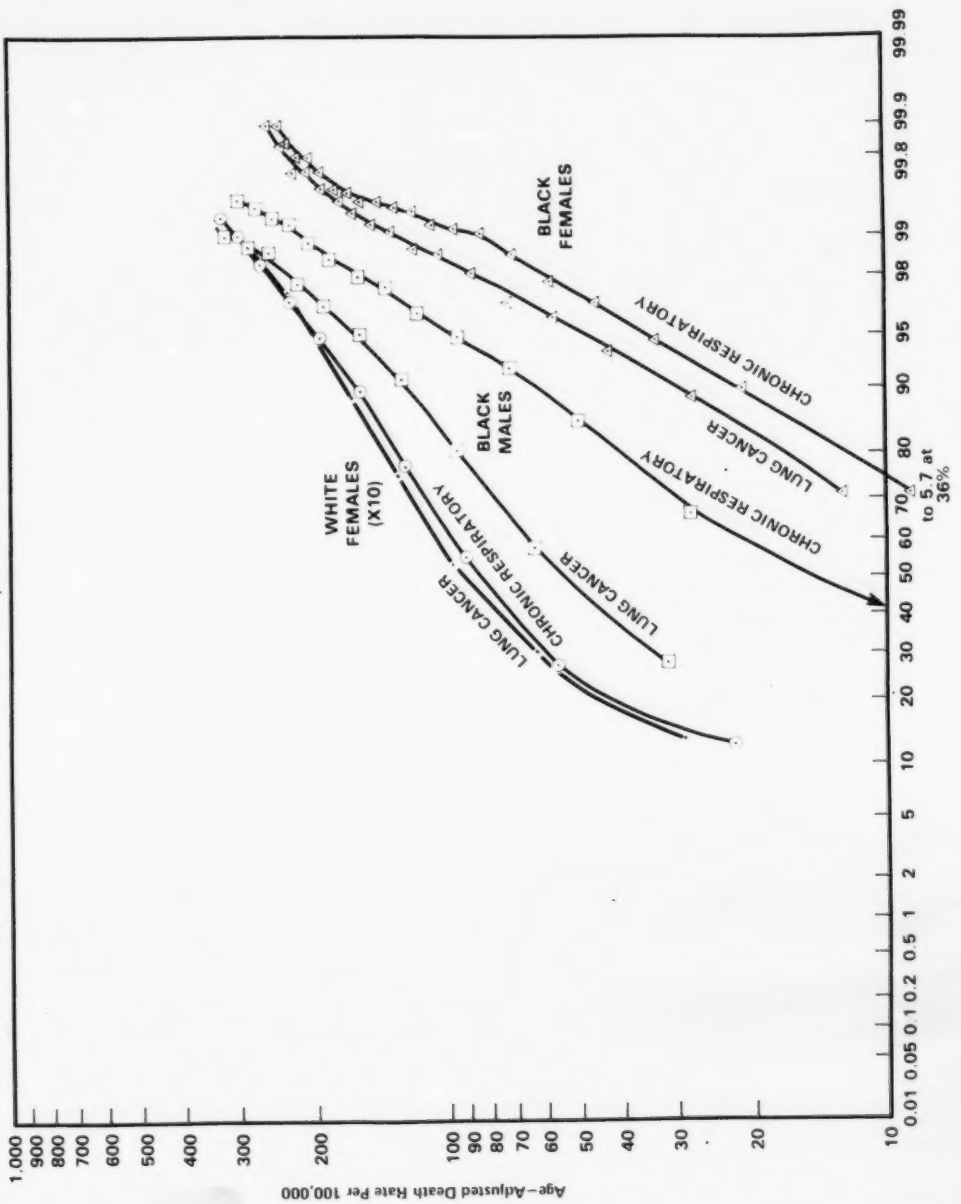
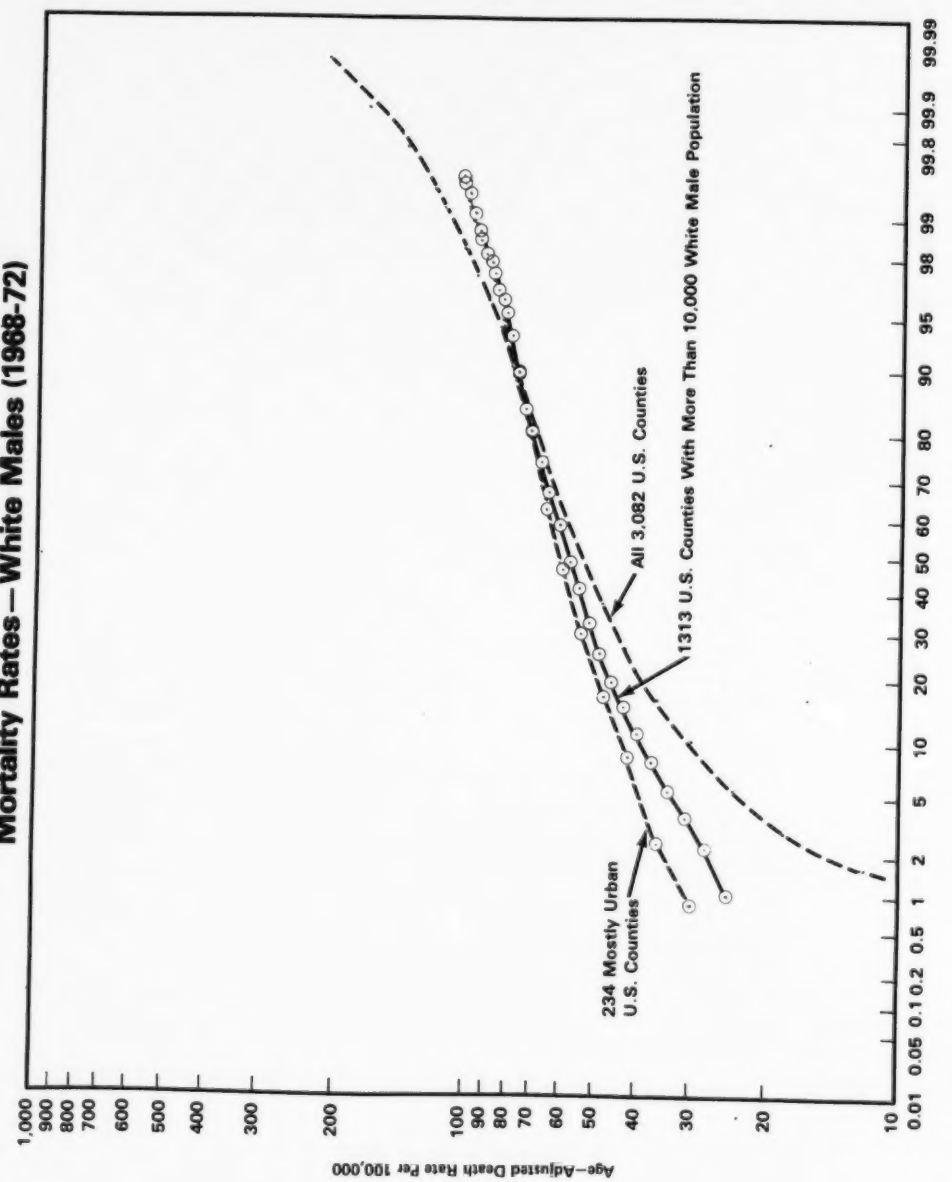


Figure 7
Cumulative Frequency Distribution of Lung Cancer Mortality Rates—White Males (1968-72)



APPENDIX A: UPGRADE Variables Used in This Study and Sauer Study with Corresponding ICDA Codes

UPGRADE CODE	VARIABLE	IDCA CODES (8th Revision)
071	Tuberculosis, All Forms	010-019
072	Other Infective Disease	000-009, 020-136
073	Ca Buccal Cavity, Pharynx	140-149
074	Cancer of Esophagus	150
075	Cancer of Stomach	151
076	Cancer of Intestine	152, 153
077	Cancer of Rectum	154
078	Ca Liver, Gall B., Ducts	155, 156
079	Cancer of Pancreas	157
080	Other Digestive Cancer	158, 159
081	Cancer of Resp. System	160-163
082	Cancer of Breast	174
083	Cancer of Cervix	180
084	Cancer of Uterus	181, 182
085	Ca Prost, Other Female Ca	183, 184, 185
086	Cancer of Bladder	188
087	Cancer of Kidney, Etc	189
088	Cancer of Central Nervous System (CNS)	191, 192
089	Residual Cancer	170-3, 183, 186-7, 190, 194
090	Cancer, Ill-Def. & Sec.	195-199
091	Lymphosarcoma, Etc	200
092	Hodgkin's Disease	201
093	Multiple Myeloma	203
094	Leukemia	204-207
095	Other Lymphatic	202, 208, 209
096	Neoplasms, Benign & Unspecified	210-239

UPGRADE CODE	VARIABLE	ICDA CODES (8th Revision)
097	Diabetes	250
098	Alcoholism	303
099	Rheumatic Heart Dis.	390-398
100	Hypertension	400-404
101	Acute Ischemic Heart Dis.	410,411
102	Chronic Ischemic Heart	412,413
103	Other Heart Disease	420-429
104	Cerebrovascular Disease	430-438
105	Arteriosclerosis	440
106	Aortic Aneurysm	441
107	Other Arteries, Etc	442-448
108	Veins, Etc	450-458
109	Influenza and Pneumonia	470-486
110	Chronic Resp. Dis.	490-493,517-519
111	Cirrhosis of Liver	571
112	Chronic Nephritis, Etc	582-584
113	Infections of Kidney	590
114	Congenital Heart & Circ.	746,747
115	Other Congenital	740-745,748-759
116	Other Early Infancy	760-778
117	Symptoms, Ill-Defined	780-796
126	Major CV Diseases	390-448
127	Cancer, All Sites and Forms	140-209

APPENDIX B: Variables Used in the Wellington, MacDonald and Wolf Study with Corresponding ICDA Codes

Response Variables: State Age-Adjusted Death Rates

Code Number[a]	Site	
1950-1969[b]		ICD 1948
M/F100	All malignant neoplasms	140-205
M/F140	Buccal cavity and pharynx	140-148
M/F150	Esophagus	150
M/F151	Stomach	151
M/F153	Large intestine excluding rectum	153
M/F154	Rectum	154
M/F155	Biliary passages and liver specified as primary and unspecified	155, Unspec. Part of 156
M/F157	Pancreas	157
M/F161	Larynx	161
M/F166	Trachea, bronchus, and lung	162-163
M/F170	Breast	170
F171	Cervix uteri	171
F172	Corpus and other parts of uterus, and uterus unspecified	172-174
F175	Ovary	175
M177	Prostate	177
M/F180	Kidney	180
M/F181	Bladder and other urinary organs	181
M/F190	Malignant melanoma of skin	190
M/F191	Other malignant neoplasms of skin	191
M/F193	Brain and other parts of the nervous system	193
M/F194	Thyroid	194
M/F200	Lymphosarcoma and reticulosarcoma	200,202,205
M/F201	Hodgkin's disease	201
M/F203	Multiple myeloma	203

APPENDIX B CONTINUED

Code Number[a]	Site	
M/F204	Leukemia and aleukemia	204
1950-1967[c]		
M/F162	Trachea, and bronchus and lung specified as primary	162
M/F163	Bronchus and lung unspecified as to primary or secondary	163
1949-1959[d]		
M/F100B	All malignant neoplasm	140-205
M/F140B	Buccal cavity and pharynx	140-148
M/F150B	Esophagus	150
M/F151B	Stomach	151
M/F153B	Intestines excluding rectum	152-153
M/F154B	Rectum	154
M/F158B	Biliary passages and liver	155-156
M/F155B	Biliary passages and liver specified as primary	155
M/F156B	Biliary passages and liver secondary and Unspecified	156
M/F157B	Pancreas	157
M/F161B	Larynx	161
M/F166B	Trachea, bronchus, and lung	162-163
M/F162B	Trachea, and bronchus, and lung specified as primary	162
M/F163B	Bronchus and lung unspecified as to primary or secondary	163
M/F170B	Breast	170
F171B	Cervix uteri	171
F172B	Corpus and other parts of uterus, and uterus unspecified	172-174
F175B	Ovary	175
M177B	Prostate	177
M/F180B	Kidney	180
M/F181B	Bladder and other urinary organs	181
M/F189B	Skin including melanoma	190-191
M/F193B	Brain and other parts of the nervous system	193
M/F194B	Thyroid	194

APPENDIX B CONTINUED

Code Number[a]	Site	
M/F200B	Lymphosarcoma and reticulosarcoma and other forms of lymphoma	200-203,205
M/F201B	Hodgkin's Disease	201
M/F204B	Leukemia and aleukemia	204
1960[e]		ICD 1955
M/F256	Gallbladder	155.1
Major Disease Categories		
1949-1959[f]		ICD 1948
M/F010B	Tuberculosis	001-019
M/F050B	Infectious and parasitic except influenza and tuberculosis	020-138
M/F400B	Circulatory system	400-468
M/F470B	Respiratory system except influenza and asthma	470-475, 490-527
M/F480B	Influenza	480-483
M/F500B	Genito-urinary system	590-637
M/F530B	Digestive system	530-587
M/F800B	Accidental and violent causes	E800-E999
M/F900B	All deaths	001-E999

[a]M, white male. F, white female.

[b]Mason, T. J., and McKay, F. W. "U. S. Cancer Mortality by County: 1950-1969." DHEW Publ. No. (NIH) 74-615, USGPO, Washington, D.C., 1973.

[c]Burbank, F."Patterns in cancer mortality in the United States: 1950-1967." Nat. Cancer Inst. Monogr. 33: 199-216, 1971.

[d]U.S. Public Health Service. "Vital Statistics of the United States," Vols. II and III, Mortality Data, 1949-1959, USGPO, Washington, D.C., 1951-1962.

[e]U.S. Public Health Service. "Vital Statistics of the United States," Vol. II, Mortality Data, 1960. USGPO, Washington, D.C., 1963.

[f]U.S. Public Health Service. "Vital Statistics of the United States," Vols. II and III, Mortality Data, 1949-1959, USGPO, Washington, D.C., 1951-1962.

APPENDIX C: All Correlations with $p \leq .0001$ (Males)

$r \geq .070$

TB, All Forms - Ca Buccal Cavity	.074
Alcoholism	.071
Chronic Resp.	.075
Other Infective Dis - Other Congenital	.078
Ca Buccal Cavity - Ca Esophagus	.075
Ca Kidney	.075
Ca Ill-Def. & Sec.	.074
Ca Esophagus - Ca Ill-Def. & Sec.	.073
Ca Stomach - Diabetes	.077
Ca Intestine - Rheum. Heart	.074
Aortic Aneurysm	.077
Ca Rectum - Diabetes	.071
Ca Pancreas - Acute Isch.	.077
Other Digestive Ca - Other Lymphatic	.073
Ca Res. - Influ & Pneum.	.078
Veins	.071
Symptoms, Ill-Def.	.074
Ca Bladder - Rheum. Heart	.077
Residual Ca - Ca, Ill-Def & Sec.	.071
Rheum Heart	-.078
Veins	.074
Cerebrovascular	.079
Symptoms, Ill-Def.	.072
Ca, Ill-Def. & Sec. - Multiple Myeloma	-.074
Veins	.074
Leukemia - Veins	-.079
Other Lymphatic - Diabetes	.073
Diabetes - Hypertension	.079

Arteriosclerosis	.075
Chronic Isch.	.077
Ca, All Forms	.074
Alcoholism - Veins	.071
Rheum. Heart - Arteriosclerosis	.077
Hypertension - Arteriosclerosis	.077
Chronic Nephritis	.076
Chronic Isch.	.076
Other Heart - Flu	.074
Infect. Kidney	.074
Arteriosclerosis - Other Arteries	.072
Aortic Aneurysm - Other Arteries	.073
Neoplasms - Chronic Resp.	.074
Veins - Symptoms, Ill-Def.	.075
$r \geq .080$	
TB, All Forms - Infect. of Kidney	.080
Ca, All Forms	.080
Other Infective Dis - Ca Bladder	-.082
Ca Buccal Cavity - Major CV	.081
Ca Esophagus - Ca Intestine	.086
Ca Stomach - Ca Bladder	.081
Ca Rectum - Cerebrovascular	-.082
Ca Intestine - Other Lymphatic	.089
Ca Resp. - Residual Ca	.086
Lymphasarcoma - Aortic Aneurysm	.084
Diabetes - Chronic Nephritis	.084
Chronic Isch.	.085
Rheum. Heart - Acute Isch.	-.087
Other Heart	-.086
Cerebrovascular	.080
Hypertension - Acute Isch.	.089
Alcoholism - Other Arteries	.084
Acute Isch. - Flu	.082
Infect. of Kidney	.087
Other Arteries	.080

Veins	.080
Other Early Infancy	.080
Major CV	.082
Cerebrovascular - Ca, All Forms	.086
Arteriosclerosis - Aortic Aneurysm	.089
Aortic Aneurysm - Symptoms, Ill-Def.	-.084
Flu - Infect. of Kidney	.081
Symptoms, Ill-Def.	.086
Cirrhosis - Major CV	.081
Other Arteries - Veins	.087
$r \geq .090$	
TB, All Forms - Other Infective Dis.	.099
Cirrhosis	.095
Chronic Nephritis	.094
Veins	.095
Other Infective Dis. - Flu	.090
Ca Esophagus - Ca Resp.	.094
Ca Stomach - Cirrhosis	.099
Ca, Ill-Def. & Sec. - Cirrhosis	.095
Neoplasms - Cirrhosis	.094
Ca Intestine - Flu & Pneum.	-.091
Rheum. Heart - Chronic Resp.	.093
Cirrhosis	.093
Alcoholism - Chronic Resp.	.099
Cirrhosis	.093
Ca Intestine - Leukemia	.095
Ca Resp.	.096
Ca Prost., O. Female Ca	.096
Ca Rectum - Aortic Aneurysm	.091
Other Digestive Ca - Ca, All Forms	.093
Multiple Myeloma - Ca, All Forms	.090
Diabetes - Major CV	.097
Acute Isch - Other Heart	-.097
Chronic Isch. - Chronic Resp.	.090

	Cerebrovascular	-.093
Chronic Resp. -	Major CV	.090
Arteriosclerosis -	Chronic Resp.	.095
	$\underline{r}$ = .100 - .120	
TB, All Forms -	Cerebrovascular	.105
	Major CV	.118
	Ca, Ill-Def. & Sec.	.117
Ca Buccal Cavity -	Chronic Isch.	.115
	Aortic Aneurysm	.100
Ca Esophagus -	Ca Rectum	.114
Ca Intestine -	Ca Bladder	.110
Ca Liver -	Ca, All Forms	.118
Ca Resp. -	Infect. of Kidney	.113
Hodgkins -	Ca. All Forms	.104
Rheum. Heart -	Aortic Aneurysm	.118
Acute Isch. -	Ca, All Forms	.102
Other Heart -	Aortic Aneurysm	-.101
Flu -	Major Cv	.119
Infect. of Kidney -	Symptoms, Ill-Def.	.102
	Major CV	.114
Ca, Ill-Def. & Sec. -	Acute Isch.	.105
	Cerebrovascular	.102
	$\underline{r}$ = .120 - .130	
TB, All Forms -	Symptoms, Ill Def.	.122
Ca Buccal Cavity -	Cirrhosis	.130
Ca Esophagus -	Chronic Isch.	.126
Ca Stomach -	Ca Rectum	.121
Ca Intestine -	Chronic Isch.	.128
	Other Heart	-.130
	Cirrhosis	.127
Ca Rectum -	Rheum. Heart	.130
Ca Resp. -	Ca Bladder	.122
Ca Bladder -	Chronic Isch.	.121
	Aortic Aneurysm	.128
Alcoholism -	Flu	.125

Cerebrovascular - Infect. of Kidney	.126
Chronic Resp. - Cirrhosis	.123
Ca. All Forms	.122
Symptoms, Ill-Def. - Major CV	-.124
$\underline{r}$ = 130 - 140	
TB, All Forms - Flu	.135
Ca Rectum - Symptoms, Ill-Def.	-.140
Other Lymphatic - Ca, All Forms	.132
Flu - Chronic Resp.	.132
Veins - Flu	.140
$\underline{r}$ = .140 - .150	
Ca Intestine - Symptoms, Ill-Def.	-.143
Ca Rectum - Other Heart	-.142
Ca Resp. - Cerebrovascular	.141
Ca Kidney - Ca, All Forms	.143
Aortic Aneurysm - Cirrhosis	.146
Major CV	.141
Other Arteries - Major CV	.141
$\underline{r}$ = .150 - .160	
TB, All Forms - Ca Resp.	.154
Ca Rectum - Ca Bladder	.156
Ca Resp. - Acute Isch.	.156
Ca, Ill-Def. & Sec. - Major CV	.157
Chronic Isch. - Symptoms, Ill-Def.	-.155
Rheum. Heart - Chronic Isch.	.158
$\underline{r}$ = .160 - .170	
Ca Buccal - Ca Resp.	.170
Ca CNS - Ca, All Forms	.162
Lymphosarcoma - Ca, All Forms	.163
Arteriosclerosis - Major CV	.164
Ca Rectum - Intestine	.168
Cirrhosis	.168
Aortic Aneurysm - Chronic Resp.	.166

$\underline{r}$ = .170 - .180	
Ca Resp. - Ca, Ill-Defined and Sec.	.175
Aortic Aneurysm	.173
Cirrhosis	.178
Leukemia - Ca, All Forms	.177
Aortic Aneurysm - Ca, All Forms	.178
$\underline{r}$ = .180 - .190	
Ca Esophagus - Cirrhosis	.186
Acute Isch. - Symptoms, Ill-Def.	-.181
$r \geq .200$	
Ca Buccal - Ca, All Forms	.283
Ca Esophagus - Ca, All Forms	.244
Ca Stomach - Ca, All Forms	.238
Ca Rectum - Chronic Isch.	.204
Ca, All Forms	.256
Ca Pancreas - Ca, All Forms	.231
Ca Resp. - Chronic Isch.	.220
Chronic Resp.	.201
Ca Prost, Other Female Ca - Ca, All Forms	.286
Ca Bladder - Ca, All Forms	.281
Residual Ca - Ca, All Forms	.222
Ca, Ill-Defined and Sec. - Ca, All Forms	.296
Hypertension - Major CV	.214
Chronic Isch. - Other Heart	.214
Aortic Aneurysm	.201
Cirrhosis	.222
Ca, All Forms	.217
Other Heart - Symptoms, Ill-Defined	.256
Cirrhosis - Ca, All Forms	.249
Major CV - Ca, All Forms	.227
$r > .300$	
Ca Resp. - Major CV	.302
Ca, All Forms	.666
Acute Isch. - Cerebrovascular	.307
Major CV	.728

Chronic Isch. - Major CV	.448
Cerebrovascular - Major CV	.515
Ca Intestine - Ca, All Forms	.393

APPENDIX D: All Correlations with $p \leq .0001$ (Females)

$r \geq .070$	
TB - Congenital Heart	.077
Other Infective Dis - Infection of Kidney	.077
Symptoms, Ill-Defined	.079
Ca Buccal Cavity - Ca Cervix	.073
Major CV	.077
Ca Intestine - Ca Liver, Gall B, Ducts	.073
Ca Rectum - Ca Resp. System	.076
Infection of Kidney	-.078
Ca Liver, Gall B, Ducts - Hypertension	.073
Ca Resp. System - Ca Cervix	.075
Ca Cervix - Ca, Ill-Defined & Sec.	.072
Chronic Resp.	.074
Ca Uterus - Hypertension	.077
Ca Bladder - Chronic Isch.	.079
Ca Ill-Def. & Sec. - Infection of Kidney	.073
Cerebrovascular - Influ. & Pneum.	.079
Ca Pros., Other Female Ca - Other Heart	-.075
Ca Stomach - Residual Ca	.079
$r \geq .080$	
TB, All Forms - Other Infective Dis	.085
Residual Ca	.087
Chronic Resp.	.088
Infections of Kidney	.082
Ca, All Forms	.080
Other Infective Dis - Influ. & Pneum.	.080
Ca Buccal Cavity - Ca Bladder	.089
Chronic Resp.	.080

Ca Rectum - Ca Cervix	.085
Ca Prost., Other Female Ca	.081
Symptoms, Ill-Defined	-.083
Ca Liver , Gall B., Ducts - Ca Prost., Other Female Ca	.084
Ca Kidney	.084
Ca Pancreas - Other Congenital	.089
Other Digestive Ca - Ca, All Forms	.080
Ca Resp. System - Ca Breast	.080
Acute Isch.	.087
Ca Breast - Ca Prost., Other Female Ca	.088
Infection of Kidney	-.083
Ca Cervix - Hypertension	.088
Acute Isch.	.087
Other Congenital	.080
Ca Uterus - Other Heart	.084
Ca, Ill-Def. & Sec. - Neoplasms	.081
Chronic Resp.	.082
Lymphosarcoma - Symptoms, Ill-Defined	-.089
Diabetes - Aortic Aneurysm	.083
Acute Isch. - Veins	.084
Other Heart - Aortic Aneurysm	-.086
Aortic Aneurysm - Major CV	.087
Veins, Etc - Infection of Kidney	-.087
Influ. & Pneum. - Major CV	.080
Cirrhosis - Major CV	.083
Congenital Heart - Other Congenital	.080
$\underline{r} \geq .090$	
TB, All Forms - Ca Buccal Cavity	.091
Ca Buccal Cavity - Cirrhosis	.090
Ca Intestine - Ca Uterus	.090
Rheum. Heart	.098
Major CV	.095
Ca Rectum - Cirrhosis	.094
Major CV	.094

Ca Breast - Lymphosarcoma	.095
Rheum. Heart	.098
Hypertension	.093
Cirrhosis	.098
Ca Pros., Other Female Ca - Rheum. Heart	.090
Ca, Ill-Def. & Sec. - Acute Isch.	.095
Lymphosarcoma - Other Heart	-.091
Other Lymphatic - Symptoms, Ill-Defined	-.092
Ca, All Forms	.095
Neoplasms - Ca, All Forms	.098
Rheum. Heart - Cerebrovascular	-.092
Chronic Resp.	.092
Hypertension - Chronic Nephritis	.094
Acute Isch. - Infection of Kidney	.092
Other Heart - Major CV	.097
Cerebrovascular - Other Arteries	.092
Arteriosclerosis - Ca, All Forms	.090
Chronic Nephritis - Major CV	.095
Infections of Kidney - Major CV	.098
Other Congenital - Major CV	.097
$\underline{r}$ = .100 - .120	
TB, All Forms - Ca Cervix	.110
Chronic Isch.	.105
Other Infective Dis - Other Heart	.111
Ca Buccal Cavity - Ca, All Forms	.115
Ca Esophagus - Cirrhosis	.118
Ca Stomach - Ca Rectum	.100
Ca Intestine - Chronic Isch.	.120
Ca Rectum - Ca Liver, Gall B, Ducts	.108
Other Heart	-.105
Ca Resp. System - Chronic Resp.	.118
Ca Cervix - Chronic Isch	.120
Cerebrovascular	.110
Influ. & Pneum.	.101
Infection of Kidney	.103

Ca Ill-Def. & Sec. - Other Heart	.109
Chronic Nephritis	.102
Diabetes - Hypertension	.107
Arteriosclerosis	.113
Ca, All Forms	.109
Rheum. Heart - Acute Isch.	-.118
Hypertension - Other Heart Dis	.109
Arteriosclerosis	.100
Acute Isch. - Influ. & Pneum	.118
Symptoms, Ill-Defined	-.118
Other Heart - Chronic Nephritis	.102
Veins - Major CV	.118
Ca, All Forms	.111
Influ. & Pneum. - Symptoms, Ill-Defined	.115
Chronic Resp. - Infection of Kidney	.115
Infections of Kidney - Symptoms, Ill-Defined	.103
Symptoms, Ill-Defined - Major CV	-.114
$\underline{r}$ =.121 -.130	
Ca Intestine - Ca Prost., Other Female Ca	.125
Other Heart	-.125
Symptoms, Ill-Defined	-.123
Ca Breast - Chronic Isch.	.128
Symptoms, Ill-Defined	-.122
Diabetes - Acute Isch.	.121
Chronic Isch. - Aortic Aneurysm	.130
Other Heart - Ca, All Forms	-.123
Veins - Cirrhosis	.122
Chronic Resp. - Cirrhosis	.130
Rheum. Heart - Symptoms, Ill-Defined	-.125
$\underline{r}$ =.131 -.140	
TB, All Forms - Major CV	.132
Ca Esoph. - Ca, All Forms	.139
Hodgkins Dis - Ca, All Forms	.137
Rheum. Heart - Other Heart	-.140
Infection of Kidney	-.137

Hypertension - Ca, All Forms	.134
Chronic Isch. - Symptoms Ill-Defined	.137
Cerebrovascular - Infection of Kidney	.139
Aortic Aneurysm - Ca, All Forms	.136
Symptoms, Ill-Defined - Ca, All Forms	.139
$\underline{r}$ =.141 -.150	
Ca Intestine - Ca Rectum	.145
Ca Breast - Other Heart	-.146
Diabetes - Chronic Isch.	.148
Rheum. Heart - Aortic Aneurysm	.141
Chronic Isch. - Cerebrovascular	-.141
Other Arteries - Major CV	.147
Veins - Influ. & Pneum.	.145
Influ. & Pneum - Chronic Resp.	.145
Infection of Kidney	.144
$\underline{r}$ =.151 -.160	
Ca Rectum - Ca Breast	.155
Chronic Isch.	.151
Ca, Ill-Def. & Sec. - Major CV	.153
Multiple Myeloma - Ca, All Forms	.155
Aortic Aneurysm - Cirrhosis	.158
Major CV - Ca, All Forms	.152
$\underline{r}$ = .161 - .170	
Ca Intestine - Ca Breast	.161
Rheum. Heart - Cirrhosis	.164
$\underline{r}$ = .171 - .180	
Ca Kidney - Ca, All Forms	.172
$\underline{r}$ = .181 - .190	
Ca Cervix - Major CV	.187
Ca Bladder - Ca, All Forms	.183
Ca Central Nervous System - Ca, All Forms	.188
Residual Ca - Ca, All Forms	.190
Leukemia - Ca, All Forms	.181
Rheum. Heart - Ca, All Forms	.189
Acute Isch. - Cerebrovascular	.189

$r = .191 - .200$	
Rheum. Heart - Chronic Isch.	.193
Arteriosclerosis - Major CV	.196
Cirrhosis - Ca, All Forms	.194
$r \geq .200$	
Ca Stomach - Ca, All Forms	.249
Ca Liver, Gall B, Ducts - Ca., All Forms	.204
Ca Pancreas - Ca, All Forms	.243
Ca Resp. - Cirrhosis	.206
Ca Cervix - Ca, All Forms	.239
Ca Uterus - Ca, All Forms	.264
Ca, Ill-Def. & Sec. - Ca, All Forms	.241
Lymphosarcoma - Ca, All Forms	.212
Diabetes - Major CV	.210
Hypertension - Major CV	.249
Chronic Isch. - Other Heart	-.216
Cirrhosis	.211
Ca, All Forms	.238
Other Heart - Symptoms, Ill-Defined	.203
$r \geq .300$	
Ca Intestine - Ca, All Forms	.434
Ca Rectum - Ca, All Forms	.332
Ca Resp. System - Ca, All Forms	.360
Ca Breast - Ca, All Forms	.547
Ca Prost., Other Female Ca - Ca, All Forms	.336
Acute Isch. - Major CV	.607
Chronic Isch. - Major CV	.575
Cerebrovascular - Major CV	.466

APPENDIX E: Ranked Intercorrelations from Sauer Study for SEA'S and States, 1968-72 (White Males)

Correlation	N = 508 SEA r-values	N = 50 State r-values
Stroke - Ischemic and Other Heart Disease	.63(1)	.68(7)
Cancer of Respiratory System - Major CV and Ill-Defined	.61(2)	.75(1)
Ca Resp. System - Ischemic and Other Heart	.58(3)	.75(2)
All Cancers - Ischemic and Other Heart	.57(4)	.75(3)
All Cancers - Major CV and Ill-Def.	.56(5)	.68(8)
Ca Resp. System - Cancer, Ill-Defined	.50(6)	.71(6)
Ca, Ill-Def. - Major CV and Ill-Def.	.50(7)	.65(10)
Cirrhosis of Liver - Cancer of Digestive System	.46(8)	.50(26)
Major CV - Influenza and Pneumonia	.46(9)	.36(29)
Ca Ill-Def. - Ischemic and Other Heart	.43(10)	.61(11)
All Cancers - Cirrhosis	.43(11)	.43(28)
Ischemic and Other Heart - Flu and Pneumonia	.41(12)	*
Ca Ill-Def. - Stroke	.40(13)	.44(27)
Ca Resp. System - Stroke	.39(14)	.50(24)
TB - Flu and Pneumonia	.36(15)	.68(9)
Cancer of Bladder - Cancer of Digestive	.33(16)	.75(4)
Ca Bladder - Cancer of Colo-rectal	.30(17)	.75(5)
Ischemic and Other Heart - Diabetes	.29(18)	.54(17)
Ca Ill-Def. - TB	.28(19)	.53(19)
Ca Digestive - Diabetes	.27(20)	.52(21)
Ca Colo-rectal - Diabetes	.24(21)	.54(18)
Cirrhosis - Flu and Influenza	.23(22)	.61(12)
Stroke - Rheumatic Heart	-.23(23)	-.57(15)
Ca Resp. System - Pancreas	.23(24)	.53(20)
All Cancers - Diabetes	.22(25)	.60(14)
TB - Cirrhosis	.22(26)	.55(16)

Correlation	SEA r-values	State r-values
All Cancers - Hypertension	.22(27)	.50(25)
Stroke - Hypertension	.21(28)	.52(22)
Cirrhosis - Neoplasms, Benign and Unspecified	.18(29)	.61(13)

* Not statistically significant at .01 level ($r < \pm .86$)

Source - Adapted from Sauer (6).

APPENDIX F: Ranked Intercorrelations from Sauer Study for SEA's and States, 1968-72 (White Females)

Correlation	N = 508 SEA r-values	N = 50 State r-values
Cancer of Colo-rectal - Cancer of Breast	.52(1)	.81(1)
Cirrhosis - Cancer of Respiratory System	.52(2)	.68(3)
Cancer of Digestive System - Ca Breast	.50(3)	.81(2)
All Cancers - Ischemic and Other Heart Disease	.48(4)	.67(5)
Diabetes - Major CV and Ill-Defined	.47(5)	.66(6)
Diabetes-Ischemic and Other Heart	.46(6)	.68(4)
All Cancers - Major CV	.45(7)	.59(12)
Ca Colo-rectal - Cancer of Uterus and Cervix	.44(8)	.63(7)
Stroke - Ischemic and Other Heart	.44(9)	.43(17)
Ca Ill-Defined - Major CV and Ill-Defined	.42(10)	.60(11)
Ca Cervix - Major CV	.42(11)	.49(15)
Ca Digestive - Ca, Uterus and Cervix	.41(12)	.59(13)
Cirrhosis - All Cancers	.40(13)	.44(16)
Influenza and Pneumonia - Major CV and Ill-Defined	.40(14)	.39(18)
Ca Colo-rectal - Ischemic and Other Heart	.34(15)	.55(14)
Stroke - Ca Uterus and Cervix	.32(16)	.62(9)
Hypertension - Ca Uterus and Cervix	.31(17)	.61(10)
Flu and Pneumonia - Cirrhosis	.24(18)	.62(8)

Source - Adapted from Sauer (6).

APPENDIX G: Ranked Intercorrelations from the Wellington, MacDonald and Wolf Study (White Males)

Males

	r	r^2(%)
1. Buccal Cavity, Pharynx - Larynx		91.3
2. Large Intestine - Rectum	.95	91.0
3. Rectum - Bladder	.91	82.5
4. Large Intestine - Bladder	.88	77.6
5. Esophagus - Large Intestine	.88	77.1
6. Larynx - Esophagus		74.5
7. Buccal Cavity - Esophagus	.86	74.1
8. Larynx - Trachea, Bronchus - Lung		71.0
9. Rectum - Kidney	.84	70.6
10. Esophagus - Bladder	.84	69.9
11. Esophagus - Rectum	.83	69.1
12. Trachea, Bronchus and Lung - Bronchus and Lung Unspec.		66.1
13. Large Intestine - Kidney	.78	60.3
14. Buccal Cavity - Trachea, Bronchus and Lung	.77	59.6
15. Kidney - Bladder	.76	57.9
16. Larynx - Bronchus, and Lung Unspec.		55.4
17. Melanoma of Skin - Other Skin		55.1
18. Larynx - Bladder		53.7
19. Stomach - Kidney	.72	52.1
20. Larynx - Large Intestine		52.0
21. Rectum - Lymphosarcoma	.72	51.8
22. Kidney - Other Skin		51.5
23. Kidney - Lymphosarcoma	.72	51.2
24. Buccal Cavity - Large Intestine	.71	50.0

APPENDIX H: Ranked Intercorrelations from the Wellington, MacDonald and Wolf Study (White Females)

Females

		r	r^2(%)
1.	Large Intestine - Rectum	.92	85.1
2.	Breast - Ovary		82.5
3.	Rectum - Breast	.89	79.8
4.	Large Intestine - Breast	.88	77.8
5.	Large Intestine - Ovary		64.8
6.	Lymphosarcoma - Ovary		60.0
7.	Rectum - Ovary		59.0
8.	Trachea, Bronchus and Lung - Bronchus, Lung Unspecified		58.4
9.	Lymphosarcoma - Breast	.85	55.5
10.	Rectum - Bladder	.74	54.8
11.	Melonoma of Skin - Other Skin		54.8
12.	Rectum - Hodgkins	.74	54.7
13.	Buccal Cavity - Other Skin		52.9
14.	Large Intestine - Bladder	.72	51.4

APPENDIX I: Intercorrelation Procedure for White Females and Males

Females

The 51 variables used for the intercorrelations for white females were divided into 5 groups: A,B,C,D, and E, each containing 10 variables, plus the single variable 61, cancer, all forms and sites. The variables in each group are listed below.

GROUP A

Key Code	Variable
11	Tuberculosis, All Forms
12	Other Infective Disease
13	Ca Buccal Cavity, Pharynx
14	Cancer of Esophagus
15	Cancer of Stomach
16	Cancer of Intestine
17	Cancer of Rectum
18	Ca Liver, Gall B., Ducts
19	Cancer of Pancreas
20	Other Digestive Cancer

GROUP B

Key Code	Variable
21	Cancer of Resp. System
22	Cancer of Breast
23	Cancer of Cervix
24	Cancer of Uterus
25	Ca Prost., Other Female Ca
26	Cancer of Bladder
27	Cancer of Kidney, Etc
28	Cancer of CNS
29	Residual Cancer
30	Cancer, Ill-Def. and Sec.

GROUP C

Key Code	Variable
31	Lymphosarcoma, Etc
32	Hodgkins Disease
33	Multiple Myeloma
34	Leukemia
35	Other Lymphatic
36	Neoplasms, Benign and Unspec.
37	Diabetes
38	Alcoholism
39	Rheumatic Heart Disease
40	Hypertension

GROUP D

Key Code	Variable
41	Acute Ischemic Heart Disease
42	Chronic Ischemic Heart
43	Other Heart Disease
44	Cerebrovascular Disease
45	Arteriosclerosis
46	Aortic Aneurysm
47	Other Arteries, Etc.
48	Veins, Etc.
49	Influenza and Pneumonia
50	Chronic Resp. Disease

GROUP E

Key Code	Variable
51	Cirrhosis of Liver
52	Chronic Nephritis, Etc
53	Infections of Kidney
54	Congenital Heart and Circ.
55	Other Congenital
56	Other Early Infancy
57	Symptoms, Ill-Defined
58	Cancer of Bone
59	Cancer of Thyroid
60	Major CV Diseases
61	Cancer, All Forms and Sites

The extraction procedure only allows 10 variables to be excluded at one time, as does the filtering procedure in UPGRADE. To create the 20 x 20 matrices, the following diagram was used as a guide to produce the intercorrelations of all 51 variables.

Exclusion Group vs Filtered Variables

A	B C D E 61
*B	C D E
*C	D E
*D	E

* To avoid running the single variable--61--against all of the exclusion groups, creating five 10 x 1 matrices, another variable--38-Alcoholism--was completely omitted from the intercorrelations. Prior experience told the authors that this variable was never significantly correlated with any of the other diseases, and it was therefore eliminated.

The number of counties resulting in each set of matrices varied depending on the filtered variables.

Exclusion Groups	vs Filtered Variables	Number of Counties
A	B	2782
	C	2711
	D	2759
	E	2705

Exclusion Groups	vs Filtered Variables	Number of Counties
	61	2850
B	C	2820
	D	2842
	E	2810
C	D	2255
	E	2305
D	E	2333

Males

Forty-eight variables were used to study the intercorrelations of mortality rates for white males, three less than for females due to the omission of cancer of the breast, cervix, and uterus. These variables were divided into four groups--A, B, C, and D--consisting of 10 variables and a 5th group--E--containing eight variables.

GROUP A

Key Code	Variable
16	Cancer of Intestine
17	Cancer of Rectum
21	Cancer of Resp. System
23	Cancer of Bladder
38	Acute Ischemic Heart Disease
39	Chronic Ischemic Heart
40	Other Heart Disease
41	Cerebrovascular Disease
47	Chronic Resp. Disease
48	Cirrhosis of Liver

GROUP B

Key Code	Variable
11	Tuberculosis, All Forms
12	Other Infective Disease
13	Ca Buccal Cavity, Pharynx
14	Cancer of Esophagus
15	Cancer of Stomach
18	Ca Liver, Gall B., Ducts
19	Cancer of Pancreas
20	Other Digestive Cancer
22	Ca Prost., Other Female Ca
24	Cancer of Kidney, Etc

GROUP C

Key Code	Variable
25	Cancer of CNS
26	Residual Cancer
27	Cancer, Ill-Defined and Sec.
28	Lymphosarcoma, Etc
29	Hodgkins Disease
30	Multiple Myeloma
31	Leukemia
32	Other Lymphatic
33	Neoplasms, Benign and Unspec.
34	Diabetes

GROUP D

Key Code	Variable
35	Alcoholism
36	Rheumatic Heart Disease
37	Hypertension
42	Arteriosclerosis
43	Aortic Aneurysm
44	Other Arteries, Etc
45	Veins, Etc
46	Influenza and Pneumonia
49	Chronic Nephritis, Etc.
50	Infections of Kidney

GROUP E

Key Code	Variable
51	Congenital Heart and Circ.
52	Other Congenital
53	Other Early Infancy
54	Symptoms, Ill-Defined
55	Cancer of Bone
56	Cancer of Thyroid
57	Major CV Diseases
58	Cancer, All Forms and Sites

Again, only excluding and filtering on 10 variables at a time, the following diagram indicates the procedure used to intercorrelate all 48 variables.

Exclusion Group vs Filtered Variables

A	B C D E
B	C D E
C	D E
*D	E

* Instead of doing the last extraction, D, the authors found it more time-efficient and cost effective to intercorrelate the remaining 18 variables in groups D and E through filtering. The procedure was as follows:

The two groups were first divided in half with five variables in each section for group D and four in each section in E.

Variable Key Code		Variable Key Code	
35 36 37 42 43	D_1	51 52 53 54	E_1
44 45 46	D_2	55 56 57	E_2

Then, each variable was filtered separately and intercorrelated as follows:

D_1 vs E_1 E_2 $\qquad$ D_2 vs E_1 E_2

The number of resulting counties was as follows:

Exclusion Group	vs Filtered Variables	Number of Counties
A	B	2896
	C	2899
	D	2862
	E	2897
B	C	2889
	D	2857
	E	2880
C	D	2857
	E	2879
D	E_1	2886
	E_2	2856
D	E_1	2879
	E_2	2851

It should be noted here that the above procedure resulted in a number of repeated correlations, due to the fact that those particular pair of variables appeared in the same exclusion group. As a result, differing r values would appear due to the varying number of counties involved in the particular correlation. Whenever this pehnomenon occurred, the r values and the number of corresponding counties were averaged together. For example:

For males, the specific correlation between variable 38--acute ischemic heart disease, and 41--cerebrovascular disease was intercorrelated four times as follows:

Intercorrelations	r value	Number of Counties
A_{36} vs B_{41}	.307	2896
C_{41}	.309	2899
D_{41}	.301	2862
E_{41}	.286	2897

Thus, the r value and number of counties averaged out to .301 and 2889 respectively. This procedure was applied to all like situations in the study.